

The Syntax of Codeswitching

Analysing Moroccan Arabic/Dutch Conversations

Een wetenschappelijke proeve op het gebied van de Letteren

Proefschrift

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit van Nijmegen,
volgens het besluit van het College van Decanen
in het openbaar te verdedigen op
vrijdag 3 juli 1998
des namiddags om 1.30 uur precies

door

Louis Patrick Clémens Boumans

geboren 22 maart 1967 te Tunis

Tilburg University Press

Promotores: Prof. Dr. C.H.M. Versteegh
Prof. Dr. C.L.J. de Bot
Co-Promotor: Dr. J.M. Nortier (UU)

Manuscriptcommissie:
Prof. Dr. P.C. Muysken (UvA)
Dr. H.W.M. Giesbers
Dr. J.J. de Ruiter (KUB)

Contents

ACKNOWLEDGEMENTS	1
PREFACE	3

PART I

A MONOLINGUAL STRUCTURE APPROACH

CHAPTER 1: HISTORICAL OUTLINE	7
1.1 Historical outline	8
1.2 Sociolinguistic approaches	10
1.3 Grammatical approaches	11
1.3.1 Linear approaches	13
1.3.1.1 Poplack's constraints	14
1.3.1.2 General objections to the linear approach	17
1.3.2 Government approaches	20
1.3.2.1 Di Sciullo, Muysken & Singh's Government Constraint	20
1.3.2.2 Belazi, Rubin & Toribio's Functional Head Constraint	21
1.3.2.3 Government approaches that include insertion	22
1.3.3 Insertion models	24
1.3.3.1 Hasselmo's Ordered Selection	25
1.3.3.2 Bautista's model of bilingual competence	30
1.3.3.3 Matrix language approaches	34
1.3.3.4 Joshi's proposal	35
1.3.3.5 Myers-Scotton's Matrix Language Frame model	35
1.3.3.6 Selectional properties of the ML (Schmid 1986)	45
Some general remarks on matrix language models	46
1.3.4 Conclusion	47
1.4 The codeswitching versus borrowing discussion	50
1.4.1 Borrowing, codeswitching, and bilingualism	52
1.4.2 Morphological integration	52
1.4.3 The time depth: from borrowing to codeswitching	54
1.4.4 Frequency of foreign lexemes	56
1.4.5 Concluding remarks	58

3.2.3.3 Dutch subordinate conjunctions in other languages	121
3.2.3.4 Malay conjunctions in Tidore	122
3.2.3.5 Spanish conjunctions in Quechua and Media Lengua	122
3.2.3.6 Conjunctions: Summary	123
3.2.4 Clause-external foregrounding strategies	123
3.2.4.1 Topic-Comment structures in Japanese/English	124
3.2.4.2 Discourse emphatic pronouns in CS with Arabic (I)	126
3.2.5 Discourse grammar: recapitulation	132
3.3 A syntactic representation of discourse grammar	133
3.3.1 A syntactic analysis in GB tradition	133
3.3.2 Primary evaluation of the CP analysis	135
3.3.3 Complications concerning the CP analysis	137
3.3.3.1 Discourse emphatic pronouns in Arabic (II)	137
3.3.3.2 Topic-Comment structures	139
3.3.3.3 Adverbs	139
3.3.3.4 Conjunctions	141
3.3.4 Summary CP analysis	141
3.4 Summary: a separate level for discourse organisation?	145

PART II

DESCRIPTION OF MOROCCAN ARABIC/DUTCH CODESWITCHING

INTRODUCTION TO PART II	153
ON QUANTIFICATION	154
CHAPTER 4: THE NIJMEGEN DATA CORPUS	157
4.1 The Moroccan community in the Netherlands	157
4.2 The Nijmegen data corpus	159
4.3 The respondents as a group	160
4.4 The individual respondents	167
CHAPTER 5: DUTCH NOUNS, NPS AND ADJECTIVES IN MA	175
5.1 Insertion of nouns	175
5.1.1 Grammatical gender	176
5.1.2 Collective nouns	178
5.1.3 Number	180
5.1.4 Definiteness	181
5.1.5 Omission of the definite prefix	187
5.1.6 Quantification	191
5.1.7 Possessive marking	196
5.1.8 Adjunct PPs	197
5.1.9 Noun complements	198

5.1.10	Attributive adjectives	200
5.1.11	Relative clauses	202
5.1.12	Interrogative forms	202
5.2	Adjective-Noun combinations	203
5.3	Insertion of adjectives	205
5.3.1	Attributive adjectives	205
5.3.2	Predicative adjectives	207
5.4	Insertion of Nominal Constituents	210
5.4.1	Nouns determined by a definite or indefinite article	211
5.4.2	Nouns determined by a possessive pronoun	212
5.4.3	Nouns determined by a quantifier	213
5.4.4	Pronominals	214
5.4.5	NPs in copula constructions	215
5.5	Grammatical functions of EL nouns and nominal constituents	216
5.6	Omission of the preposition <i>l</i>	218
5.7	Summary	220
CHAPTER 6: DUTCH VERBS AND THEIR COMPLEMENTS IN MA		223
6.1	Embedded verbs without complement	226
6.2	Pronominal complements of embedded verbs	228
6.2.1	MA Indirect Object suffixes	228
6.2.2	MA Direct Object suffixes	231
6.2.3	Omission of Object suffixes	232
6.2.4	MA pronoun suffixes in MA prepositional complements	234
6.2.5	Dutch personal pronouns and double marking	235
6.2.6	Pronouns in Dutch prepositional complements	236
6.2.7	Reflexive pronouns	238
6.2.8	Indefinite and demonstrative pronouns	240
6.3	Lexical complements of embedded verbs	242
6.3.1	Word order	242
6.3.2	Dutch lexical Direct Object complements	245
6.3.3	MA lexical Direct Object complements	248
6.3.4	Dutch lexical prepositional complements	249
6.3.5	MA lexical prepositional complements	251
6.3.6	Predicative complements	252
6.3.7	Complements of auxiliary verbs	253
6.4	Clausal complements of embedded verbs	253
6.4.1	Dutch clausal complements	253
6.4.2	MA clausal complements	254
6.5	Adverbial modification of embedded Verbs	255
6.6	Minor patterns of verb insertion	259
6.7	Consistency and variation among the respondents	261
6.8	Grammaticalisation	262
6.9	Complementation patterns	264

6.10 Evaluation in the light of the MSA	267
CHAPTER 7: DUTCH PREPOSITIONAL AND ADVERBIAL CONSTITUENTS IN MA	271
7.1 Embedded PPs	271
7.1.1 Complement PPs	271
7.1.2 Non-modal adjunct PPs	273
7.1.3 Modal and conjunctive adjunct PPs	277
7.1.4 PPs as predicates	278
7.2 Embedded Adverbs and AdvPs	279
7.2.1 Place and time adverbs	281
7.2.2 Adverbs of frequency and degree	281
7.2.3 Manner adverbs	282
7.2.4 Modal and conjunctive adverbs	284
7.3 Summary	286
CHAPTER 8: DUTCH CLAUSES AND DISCOURSE MARKERS IN MA	289
8.1 Embedded clauses	289
8.1.1 Relative clauses	289
8.1.2 Conditional clauses	291
8.1.3 Embedded Dutch complement clauses	292
8.1.4 Reported speech	294
8.1.5 Adverbial clauses	295
8.2 Discourse markers	296
8.2.1 Conjunctions	297
8.2.2 <i>ja</i> and <i>nee</i>	297
8.2.3 The question tag <i>hè</i>	298
8.3 Summary	299
CHAPTER 9: MA INSERTIONS IN DUTCH	301
9.1 MA nouns, nominal constituents and adjectives in Dutch matrices	301
9.1.1 Insertion of nouns and adjectives	301
9.1.1.1 Moroccan and Islamic terminology	302
9.1.1.2 Embedded MA words in repetitions	304
9.1.2 Insertion of nominal constituents	307
9.1.2.1 Nouns determined by an article	307
9.1.2.2 Topic NPs in copula clauses	310
9.1.2.3 Topic NPs in other than copula clauses	312
9.1.2.4 Left-dislocated NPs, or clause-external Topics	313
9.1.3 Summary	313
9.2 MA Prepositional and Adverbial Constituents in Dutch matrices	315
9.2.1 Embedded PPs	315
9.2.2 Embedded Adverbs and AdvPs	317
9.2.2.1 Place and time adverbs	317
9.2.2.2 The words <i>matalen</i> and <i>bhal daba</i> “for example”	319

9.2.3 Summary	320
9.3 MA Clauses in Dutch Matrices and Discourse Markers	320
9.3.1 Embedded clauses	320
9.3.1.1 Relative clauses	322
9.3.1.2 Conditional clauses and temporal adjuncts	322
9.3.1.3 Subject clauses	324
9.3.1.4 Complement clauses	325
9.3.2 MA conjunctions and other discourse markers	327
9.3.2.1 Causal conjunctions	328
9.3.2.2 Adversative conjunctions	329
9.3.2.3 Coordinate conjunctions	330
9.3.2.4 The interrogative particle <i>was</i>	331
9.3.3 Summary	332
9.4 Individual variation among respondents	332
9.5 Overview table	334

PART III

CONCLUSIONS AND OUTLOOKS

CHAPTER 10: EVALUATION OF THE MSA	339
10.1 Summary of the corpus description	339
10.1.1 Embedded nouns	340
10.1.2 Embedded adjectives	341
10.1.3 Embedded nominal constituents	341
10.1.4 Embedded verbs	341
10.1.5 Embedded prepositional constituents	342
10.1.6 Embedded adverbial constituents	343
10.1.7 Embedded clauses	343
10.1.8 Discourse marking	343
10.2 Evaluation of the Monolingual Structure Approach	343
10.2.1 Supportive evidence	345
10.2.1.1 The definition of the ML on the finite clause level	346
10.2.1.2 The definition of the ML on the constituent level and layered insertion	347
10.2.1.3 General assessment	349
10.2.2 Challenges	350
10.2.2.1 Impact of embedded verbs	350
10.2.2.2 Bare forms and double marking	354
10.2.2.3 Attributive adjectives	355
10.2.3 Discourse markers	356
10.2.4 Summary	357
10.3 How many grammars are involved in CS?	358

10.3.1 Modularity	358
10.3.2 A grammar for the CS variety	359
10.4 Patterns of codeswitching	361
10.4.1 Asymmetric Roles of Moroccan Arabic and Dutch	361
10.4.2 Co-occurrence Restrictions	362
CHAPTER 11: EXPLAINING PATTERNS OF INSERTION	365
11.1 Asymmetry in codeswitching	366
11.1.1 Community Language versus Superimposed Language	367
11.1.2 CS patterns and types of bilingual communities	369
11.2 Community Language insertions	373
11.2.1 The Matrix Language Turnover	374
11.2.2 Community Language insertions	375
11.3 The Modification and Complementation Restrictions	385
11.3.1 Collocational ties	386
11.3.2 The Modification Restrictions	387
11.3.3 The Complementation Restriction	390
11.3.4 Disappearance of co-occurrence restrictions	392
11.4 Repetition	393
11.4.1 Textual cohesion	393
11.4.2 Counter-examples to the modification restrictions	394
11.4.3 Layered insertion	397
11.5 Summary	398
OUTLOOKS	401
NOTATIONAL CONVENTIONS	403
In the main text	403
In numbered examples	404
Transcription of Moroccan Arabic	405
REFERENCES	409
APPENDIX	425
SAMENVATTING	427
CURRICULUM VITAE	433

Acknowledgements

The research on which this thesis is based was carried out at the Department of Languages and Cultures of the Middle East of the University of Nijmegen. It formed part of a larger project called *Zwaartepunt Taal en Minderheden* which included seven studies on Moroccan Arabic and Turkish in the Netherlands, supported by the Netherlands organisation for scientific research NWO and carried out at the universities of Tilburg and Nijmegen.

I wish to thank my colleagues at the Department of Languages and Cultures of the Middle East in Nijmegen and the Research Group Language and Minorities in Tilburg for their inspiration and help. A special word of thanks goes to Marijke Post in Nijmegen for her warm hospitality and to my Tilburg colleagues who wholly or partially organised many of the conferences in which I participated, both in Tilburg and in Morocco. A further thank you goes to the *Centre d'Etudes et de Recherches sur le Berbère et l'Arabe Maghrébin* (CERBAM) at the INALCO in Paris for inviting me to their regular Arabic dialectology workshops.

I am indebted to the following persons for sharing with me their knowledge and views on Moroccan Arabic, codeswitching, bilingualism or linguistics in general, for commenting on my work and for sending me copies of articles and books: Farida Abu Haidar, Ad Backus, Rob Beekes, Elizabeth Bergman, Petra Bos, Dominique Caubet, Abderrahman El Aissati, Mohamed Embarki, Bill Foley, Jan Hoogland, Maarten Kossmann, Mena Lafkioui, Carol Myers-Scotton, Jonathan Owens, Vincent de Rooij, Jan Jaap de Ruiter and Robert Vigneault. Miriam van Staden made many valuable comments on a draft version of Chapter 3; these are incorporated in the present version. Needless to say, I alone am responsible for the remaining flaws and the views expressed in this chapter, as elsewhere. I also want to thank my friends Jean Harnois and Alain Messaoudi whose hospitality I enjoyed during my visits to Paris.

There is no linguistic research without language data. I thank Youssef Azghari and my predecessor on the codeswitching project, Jacomine Nortier, for collecting and organising the Nijmegen text corpus of Moroccan Arabic/Dutch conversations. Thank you to the anonymous respondents who participated in the recording sessions. Listening to your discussions three years after the recordings was stimulating for more than linguistic reasons alone.

Like most people I owe much to my family and friends. I especially thank Pieter Jan Nellestijn who, in a way, belongs to both categories. In addition to being cheerful company, he was my personal help desk who solved many of the hardware and software problems that plagued me and my home computer. My parents enabled me

to study and always supported me. This thesis is dedicated to the memory of my mother, Elly Boumans-Berkemeijer (1930-1996).

Utrecht, February 1998.

Preface

The study of codeswitching has attracted a lot of attention during the past decades. Numerous studies deal with language choice and the sociolinguistic, psycholinguistic and grammatical properties of bilingual conversations. The present study focuses on the morphological and syntactic aspects. Today there is hardly any controversy concerning the fact that the grammatical features of codeswitching are rule-governed. However, the nature of these rules and the way in which they should be described and analysed remain a matter of much debate. With respect to the regularities observed in codeswitching within sentences, insertion models have become the most influential approach; this is due particularly to Myers-Scotton's authoritative publications since the early 1990s. Insertion models view codeswitching as the insertion of elements from one language (the embedded language) into sentences or constituents which are built according to the rules of another language (the matrix language). For this type of approach, the crucial questions are how the syntactic frame, or matrix, can be identified and what types of elements may be inserted.

The first aim of this study is to contribute to the debate on how mixed sentences can be analysed and interpreted from a grammatical point of view. The various approaches will be reviewed and an insertion type model combining insights from a number of earlier models will be proposed. The central idea is that a uniform descriptive model of codeswitching is needed in order to compare the grammatical regularities in different data sets including different language pairs and different sociolinguistic settings. An insertion model is best suited for this purpose. Even if some data appear to undermine this model, these data can be uniformly and economically described if a matrix language is assumed. For the investigation of the relationship between grammatical characteristics of codeswitching and sociolinguistic features of the language contact situation it is imperative that the grammatical analysis occurs independently of any sociolinguistic considerations. Accordingly the matrix and embedded languages must be identified solely on the basis of morphological and syntactic criteria.

The assumption of a matrix language constitutes in itself the first level of explanation of the regularities in codeswitching since a large part of these regularities is ascribed to the rules of the matrix language. Once the matrix language has been defined, patterns of embedded language insertions can be investigated. Different explanatory concepts can be advanced for these patterns on a subsequent level of explanation. At this level the comparison of uniformly described sets of codeswitching data is a valuable method.

The second aim is to provide a detailed description of Moroccan Arabic/Dutch codeswitching as spoken in the Netherlands. I will present an inventory of the syntactic and morphological regularities found in a corpus of audio-recorded conversations in Moroccan Arabic and Dutch which was gathered at the University of Nijmegen in 1990 and 1991. As this text corpus is highly heterogeneous I will point out both the idiolectal and shared features of the respondents' bilingual behaviour.

This thesis is organised as follows. The eleven chapters are divided into three main parts. Part I (Chapters 1-3) approaches codeswitching from a theoretical and cross-linguistic perspective. Chapter 1 gives an overview of the literature and reviews the major approaches to the grammar of codeswitching. Chapter 2 advances a new insertion model which combines the merits of previous proposals. This model will be called the Monolingual Structure Approach. Chapter 3 discusses various types of data that challenge this approach or cannot be satisfactorily dealt with within the insertion paradigm in general; it will be argued that much of these problematic data belongs to the realm of discourse grammar.

Part II is entirely devoted to the description of Moroccan Arabic/Dutch. The corpus description is segmented into six chapters: Chapter 4 presents a characterisation of the text corpus as a whole, its history, the individual respondents who participated in the recordings and the Moroccan community in the Netherlands. Chapters 5 to 8 discuss the insertion of Dutch elements in Moroccan Arabic clauses or sub-clausal constituents. Moroccan Arabic insertions in Dutch clauses or constituents are dealt with in Chapter 9. This final chapter of the descriptive part also includes a note on the idiolectal variation among the respondents and an overview table of insertion types.

Part III consists of two chapters. In Chapter 10 the Monolingual structure Approach is evaluated in the light of the Moroccan Arabic/Dutch data as described in part II. Finally, Chapter 11 addresses the explanation of patterns of insertion and presents proposals and avenues for further research in an explorative manner.

Part I

A Monolingual Structure Approach

A fairly reliable way to separate Ten-spot from Two-spot varieties is by leg colour. They are orange in Ten-spot Ladybirds and black in Two-spot Ladybirds.

Jiří Zahradník (1991)
The Illustrated Book of Insects p. 154.

Chapter 1

Historical Outline

The study of codeswitching has become a flourishing branch of research in the last 25 years, receiving attention from such diverse fields as sociolinguistics, anthropological linguistics, language teaching, language contact, formal linguistics and psycholinguistics, as well as from specialists in the various languages involved. Especially during the last decade so many articles and dissertations have appeared, that giving a complete review of all the literature, or even a complete bibliography, would be close to impossible. Besides, there are already some excellent overviews of the codeswitching literature (Giesbers, 1989; Haust, 1993; Myers-Scotton, 1993b, 1997; Muysken, 1995). Consequently I will only give a brief outline of the history of this field of research and focus more specifically on the study of grammatical properties of codeswitching.

In doing so, I will demonstrate a clear preference for the insertional, as opposed to the alternational or linear approach. The linear approach presents an inventory of the possible points of juxtaposition of elements of either language, e.g. a switch can be found to occur between a noun from language A and a relative pronoun from language B, between a noun and an adverb, and so forth. Restrictions on codeswitching, that is, logically possible juxtapositions that are not attested, are explained by differences in word order between the two languages. The insertional approach, on the other hand, views codeswitching as the embedding of elements from one language into a syntactic frame from the other language, the matrix language. In this case, restrictions on possible codeswitching patterns are attributed to the grammatical properties of the matrix language and to constraints on the kinds of elements that may be embedded.

In the course of this chapter I will point out the considerations leading to my preference for the latter approach for the analysis of syntactic and morphological aspects. I will concentrate on three observations that emerge as being most crucial to the analysis of codeswitching: firstly, the acknowledgment that the two languages involved play unequal roles, i.e. the matrix versus embedded language distinction; secondly, the different status of content words on the one hand, and inflection and function words on the other; and finally, the importance of congruence or structural equivalence between categories of both languages. These three observations are in fact the indispensable components of any insertion model of codeswitching: a) insertion presupposes a frame (matrix) in one language in which to insert elements from the other; b) this frame is basically a grammatical structure marked by word order, inflections and function words; and c) for an embedded language element to

fit into a slot in a matrix language frame it must somehow be perceived by the speaker as congruent to the matrix language element that would otherwise fill this slot. The asymmetric roles of the two languages involved, the content and function word dichotomy and congruence are also viewed as factors explaining regularities found in codeswitching data. Congruence only been receiving considerable attention as an independent factor in recent years.

This chapter begins with a historical outline of codeswitching studies in general (section 1), followed by a brief discussion of sociolinguistic studies in section 2. After that, section 3 addresses three types of grammatical approaches to codeswitching, namely linear, government and insertion approaches. Those approaches that have been the most fruitful and which, therefore, form the basis of this thesis will be dealt with separately and in more detail in section 3.3. The final section of this chapter presents my view on the differentiation of codeswitching and borrowing.

1.1 Historical outline

The study of what we now call ‘codeswitching’ is generally assumed to have begun with the writings of Haugen (1953: 65), and Weinreich (1953: 73) in the early fifties. There are some earlier explicit accounts on codeswitching (Schuchardt, 1884: 9 cited in Rosetti 1945/49: 77 n.1; Espinosa 1917; Rottenberg 1937), however most writing on so-called *Sprachmischung* or mixed languages actually deals with conventional lexical or morphological borrowing (e.g. Deeters, 1926; Fausel, 1959, 1962). On the other hand, the distinction between borrowing and codeswitching is not clear-cut, and some earlier writings treat instances of ‘spontaneous borrowing’ that would probably fall within the definition of codeswitching (e.g. Pap, 1949; Shanmugan Pillai, 1968).

It was Haugen (1956, 1958, 1973) who, after some conceptual confusion, coined the term *CODE SWITCHING* for “the alternate use of two languages” (1956: 40) “including everything from the introduction of a single, unassimilated word up to a complete sentence or more into the context of another language” (1973: 521). This definition has remained predominant ever since.

There has been some variation in the use of the term, especially at the ‘edges’ of the phenomenon: some authors (notably Gumperz) include switching languages as a result of a change in speech situation, thus coming close to *LANGUAGE CHOICE*; Deuchar (1995) actually confines the term codeswitching to language choice, so that there can be no codeswitching in an utterance unless the bilingual speaker is able to express the same content in both of his languages. Others exclude switching of single words, defining these as borrowings, or ‘nonce borrowings’ (notably Poplack, cf. sections 3.1 and 4 below). Further-more, some authors have tried to promote alternative terminology (e.g. code alternation, code change, code mixing, code shift, language mixture etcetera), or propose to use one term (often code-switching) for switches between sentences, and another (code-mixing) for switching within

sentences. I will stick to the term codeswitching as defined by Haugen (1973), because this term has become the most widely used and understood.

The first studies dealing with codeswitching as a primary topic appeared in the 1960s (Diebold, 1963; Gumperz, 1964; Lehtinen, 1966; Clyne, 1967). Two main branches of codeswitching research can be distinguished from the very beginning: sociolinguistic studies of the communicative function of bilingual discourse and narrower linguistic studies of grammatical aspects.

The research field as a whole received a particular boost in the US after the reclamation of linguistic rights for minorities resulted in the proclamation of the Bilingual Education Act in 1968 (Hasselmo, 1980: 157). Codeswitching studies formed a substantial part of the numerous studies on bilingualism, language use and ethnicity, many of which dealt with Spanish speaking Chicanos or Puerto Ricans. Bilingual schooling was initially intended as a transitional institution to overcome the problem of poor educational achievements in certain ethnic groups. Well aware of this context, researchers were eager to avoid further stigmatization of disadvantaged communities, for codeswitching was (and still is) generally associated with a lack of linguistic competence and an inability to speak any language properly. As a result there was a motivation to study codeswitching as an additional type of bilingual competence that was both meaningful and constrained by grammatical rules. It was part of a larger overall tendency in the 1960s and 1970s to reappraise non-standard varieties of linguistic behaviour along the lines of Labov.¹

The Spanish/English studies were the initial forum for the formulation of so-called CONSTRAINTS, i.e., syntactic sites where switching could not occur. The formulation of constraints was the 1970s' reaction to the then prevalent belief that codeswitching occurred at random and that speakers changed from one language to another at whim. With the tendency for these constraints to claim universal value and to be considered valid for any pair of languages, the interest in codeswitching with other languages than Spanish/English increased. This resulted in a multitude of articles and theses validating or invalidating the various proposed constraints for particular language pairs. Moreover, we now have at our disposal descriptions of codeswitching in a variety of language contact situations all over the world, mostly involving either English or French as prestige and educational languages in Africa and Asia, or immigrant or indigenous minority languages in North America, Europe and Australia and New Zealand, or bilingual communities on the linguistic borders in Europe (Strasbourg, Brussels, German/Danish border). Though the number of language pairs discussed is large and still increasing, most studies involve at least one of the major European languages, often English or French.

¹ Though the appreciation of codeswitching as a form of linguistic competence was dominant for the last 25 years, some authors now express less favourable judgments. It is argued that the usage of lexical items from the culturally dominant language impairs the acquisition and development of native language vocabulary (Drapeau, 1995; Sounkalo, 1995).

1.2 Sociolinguistic approaches

Sociolinguistic approaches (e.g. Gumperz, 1964 and later articles; Gumperz & Hernández-Chavez, 1971; Blom & Gumperz, 1972; Auer, 1984; various contributions in Heller (ed.), 1988; Myers-Scotton, 1993a) investigate ‘who speaks what language to whom, and when’ (Fishman, 1965). The emphasis is on the speech event and the factors that motivate switches. Techniques of discourse analysis are applied in order to attribute social meaning to codeswitching as a speech style or to individual instances of codeswitching. Blom & Gumperz (1970), for instance, distinguish between SITUATIONAL SWITCHING, which is an adaptation to a change in the speech situation and METAPHORICAL SWITCHING, which as a stylistic device can convey various social meanings. Changes in the speech situation include the arrival of a new interlocutor and shifting to another conversation topic. Metaphorical switching typically refers to a basic “we/they” social dichotomy that is reflected in the associated languages. Myers-Scotton (1988), to cite another example of the sociolinguistic approach, differentiates between switching as a MARKED CHOICE, which involves “a negotiation to change the social distance holding in the current talk exchange” (1988: 62), and switching as an UNMARKED CHOICE, which is an expected mode of discourse and “encodes dual identities for the speaker” (1988: 62). Codeswitching in the BILINGUAL CLASSROOM has developed as a distinct genre within the sociolinguistic approach thanks to the large number of publications on this topic (e.g. Zentella, 1981, Merrit, Cleghorn, et al., 1992; Arthur, 1995). The sociolinguistic approach does not primarily investigate grammatical characteristics of codeswitching, although it has stimulated research in this field. After all, any expression of social meaning presupposes regularity in language use in order to be interpretable to the listener.

Apart from the discourse-oriented approach of attributing communicative value to the act of codeswitching itself, a number of studies relate patterns of codeswitching to sociolinguistic variables, in particular LANGUAGE PROFICIENCY. Poplack (1980) found a correlation between what she called ‘intimate’ intra-sentential switching and high bilingual proficiency, as well as between ‘easy’ switching at sentence boundaries and low proficiency in either language (the terms ‘intra-’ and ‘intersentential’ will be explained in section 3). In the same vein, Nortier (1990: 113-6) reports on the relationship between bilingualism and codeswitching patterns in Moroccan Arabic/Dutch. The opposite view is held by Singh (1995): intrasententially mixed language is characteristic of weak bilinguals, whereas a high level of proficiency in both languages is required to produce inter-sentential switching. Singh’s view is supported by, and in part based on, Bentahila & Davies’s (1991, 1992) observation that switching between Moroccan Arabic and French at clause boundaries is common among their older ‘balanced’ bilingual informants, and much less so among the younger, ‘Arabic dominant’ speakers. Bentahila & Davies, however, relate this finding to the different functions the French language has for balanced and Arabic-dominant bilinguals in peer-group interaction, thus relating it to proficiency in a less direct way.

In addition to proficiency in either language, codeswitching patterns can be related to PROFICIENCY IN CODESWITCHING itself. There is evidence that learning and practising is required for the production and perception of certain speech varieties that include codeswitching, and that mere knowledge of two languages does not suffice (Hasselmo, 1969; Ward, 1975). Poplack, Wheeler & Westwood (1987), for instance, found that the ‘flagged switching’ (see below) in their Finnish/English materials was accompanied by production difficulties such as pauses and ratification markers (*uh huh, joo, niin*). They attributed this finding to the fact that their informants did not belong to a community in which either ‘nonce borrowing’ (see below) or codeswitching is a discourse mode.

Much codeswitching is found among linguistic minorities who are increasingly exposed to the economically and (hence) culturally superimposed language of their society. These bilingual communities are likely to make more use of the superimposed language over time and eventually to shift completely to that language within a few generations. This development, which is typical of immigrant communities, is reflected in the codeswitching patterns of subsequent GENERATIONS. As overall usage of the immigrant language decreases and usage of the majority language increases, codeswitching entails larger chunks of the latter and smaller contributions from the former (Torres, 1989; Pujol, 1991; Backus, 1996a,b; Backus & Boumans, 1996). According to Backus - with whom I concur - the relationship between codeswitching patterns and generations is an indirect one, which results from the relation between codeswitching patterns and the individual’s use of either language in everyday life.

A final topic related to the sociolinguistics of codeswitching is the use of DISCOURSE MARKING. From an examination of codeswitching in various settings it appears that discourse grammar is situated on a level that is distinct from sentence grammar, and that discourse markers display a distribution pattern that is different from both function morphemes and content words. Moreover, whether discourse markers and discourse marking strategies from the community language are used in the culturally superimposed language, or the other way round, seems to be strongly related to sociolinguistic factors (De Rooij, 1996: 169). While this subject certainly calls for more comparative investigation, there are a number of studies that deal with discourse markers in codeswitching contexts, e.g. Hasselmo (1970), Salmons (1990), Mashler (1995), and De Rooij (1996). In the present study, discourse marking will be addressed in detail from a syntactic point of view in Chapter 3.

1.3 Grammatical approaches

Prior to the presentation of the various grammatical analyses it may be useful to briefly mention the common distinction between switching within the same sentence and switching between sentences. These two broad categories have traditionally been called INTRA- and INTER-SENTENTIAL codeswitching. A third category, which is

sometimes labelled EXTRA-SENTENTIAL, consists of the use of various kinds of discourse markers from another language than the language of the preceding or following utterance. The three broad categories are exemplified below by passages from the Moroccan Arabic/Dutch data set gathered in Nijmegen.

- (1) ta huma xess-ek t-dir-hūm kans gev-en baš y-dir-u scor-en
 also 3PL you·must 2-do·IMPF-3PL chance give-INF so·that 3-do·IMPF-PL score-INF
 “Them also, you have to give them a chance to score.”
 Moroccan Arabic/Dutch (Abdelkrim)²

- (2) ik kijk wel daar-heen als ik tijd heb. mnin ka-t-kun ſend-i
 I look AFFIRM there-to when I time have. when ASP-3-be at-1SG
 l-weqt ka-n-šuf t-tilivizyun
 DEF-time ASP-1-look DEF-television
 “I watch that when I’ve got time. When I’ve got time I watch TV.”
 Moroccan Arabic/Dutch (Hocine)

- (3) b l-ſeks, lli mğetty-a ka-y-kun .. ka-y-kun-u l-ſeyn-in ſli-ha,
 with DEF-contrary REL covered-F ASP-3-be ASP-3-be-PL DEF-eye-PL on-3F

weet je wel?

know you AFFIRM

“Quite the reverse, the one who’s covered, all eyes will be fixed upon her, you know?” Moroccan Arabic/Dutch (Fatima)

There seems to be little controversy at this level of analysis, although the categorisation of codeswitching between coordinate and subordinate clauses as either intrasentential or intersentential is not always so straightforward. Related to this ambiguity are questions concerning the syntactic status of conjunctions, which form a heterogeneous group anyway, and which cannot always be easily distinguished from discourse markers. Obviously, the most interesting category for the study of grammatical regularities is that of intrasentential switching. It has received most attention in the literature and will also be the focus of the present study.

I follow Appel & Muysken (1987: 123-5), who divide the study of syntactic qualities of intrasentential codeswitching into a LINEAR and a STRUCTURAL approach. The linear approach describes codeswitching basically by investigating the word categories or constituents between which there can or cannot be a switch. Word order is used to explain the patterns found. This approach characterized the Spanish/English

² The names between brackets refer to the respondents who are introduced in Chapter 4. Conventions with respect to the notation in numbered examples and the transcription of Moroccan Arabic are listed at the end of this book (p. ?).

studies of the 1970s and is continued today in the work of Poplack and others following her approach. The structural approach assumes that the regularities in code-switching must be explained by some structural mechanism other than word order, notably hierarchical constituent structure. The description of codeswitching follows the lines of this ordering mechanism and is therefore also non-linear. Like most dichotomies, this is not a very strict one: many authors combine elements of both the linear and the structural approach.

Within the structural branch a number of studies adopt the concept of GOVERNMENT to explain regularities in codeswitching. Particularly, switching is predicted not to occur where certain government relations hold. Some government approaches allow for governed constituents to contain insertions from the other language; I will return to this in section 3.2 hereafter.

Before the government approaches emerged, other studies already assumed INSERTION to be the basic mechanism in codeswitching. By their recognition of insertion as the key principle, insertion models can be distinguished from the linear and government models in which insertion plays a secondary role. A logical step in the development of the insertional approach was the identification of a MATRIX LANGUAGE. Matrix language approaches, including the present work, take matrix or base language to be the main ordering force, and attribute regularities to various psycholinguistic factors such as sentence planning and the organization of the mental lexicon.

Notice that both the government and the matrix language approach assume unequal roles for the languages involved in codeswitching. At the level of an individual grammatical structure at least, one language is associated with the governor, or assumes the role of matrix in providing a grammatical frame, while elements from the other language are governed or inserted. The linear approach, on the other hand, essentially views both languages as playing equal roles, and a sequence in either language is the output of the grammatical rules of that language. Or, as Pujol (1991: 40) puts it:

Nous définissons l'AL [l'alternance de langue] comme la juxtaposition de phrases ou de fragments de phrase, cohérents et fidèles aux règles morphologiques et syntaxiques de la langue de provenance.

The following evaluation of codeswitching models is organised along the lines of this classification: Matrix language models are treated as a subset of the insertion models. Together with the government approaches, insertion models form the set of structural approaches, which are contrasted with the linear approaches.

1.3.1 Linear approaches

Amongst the earliest proposals within the linear approach was Timm's (1975) assumption of five constraints on Spanish/English codeswitching. Switching was found not to occur, a) between pronominal Subjects or objects and finite verbs; b)

between finite verbs and their infinitive complements; c) between auxiliaries and main verbs; d) between negation and negated verb; or e) in certain noun phrases containing an adjective (Timm, 1975: 477-80). In the same vein, Abbassi (1977: 157-63) formulated a number of linear constraints for Moroccan Arabic/French. The theories of universal grammar that dominated linguistic thinking instigated a quest for universal codeswitching constraints. Several authors claimed the constraints they formulated to be valid for all data corpora in any pair of languages.

1.3.1.1 Poplack's constraints

From 1978 onwards, Poplack, who initially worked on Spanish/English codeswitching, attracted much attention with her FREE MORPHEME CONSTRAINT and EQUIVALENCE CONSTRAINT. The former entailed the inhibition of codeswitching "between a bound morpheme and a lexical form unless the latter has been phonologically integrated into the language of the bound morpheme" (Sankoff & Poplack, 1981: 5). The latter predicted that "the order of sentence constituents immediately adjacent to and on both sides of the switch point must be grammatical with respect to both languages involved simultaneously" (ibid.). This means that codeswitching is prohibited elsewhere, as illustrated in (4) and (5), reproduced from Poplack, Wheeler & Westwood (1989: 132).³

- | | | |
|-----|-------------------------|---------------------|
| (4) | English: DET + N | CS: Eng DET + Sp N |
| | Spanish: DET + N | Sp DET + Eng N |
| (5) | English: ADJ + N | CS: *Eng ADJ + Sp N |
| | Spanish: N + ADJ | *Sp ADJ + Eng N |
| | | *Eng N + Sp ADJ |
| | | *Sp N + Eng ADJ |

However, these rules appeared to be only characteristic of the Spanish/English data they used, rather than universally applicable. Actually both constraints are reminiscent of earlier work on Spanish/English codeswitching: Pfaff's (1976, 1979) observations on switching adjectives and verbs, McClure's (1977: 99) and Wentz & McClure's (1977: 706-7) constraint on the 'bicodal word' and a kind of equivalence constraint that had been formulated by Lipski (1978).

While it is questionable whether the constraints are valid for Spanish/English (cf. for instance, the objections in Doron, 1983: 39, and Schmid, 1986: 71), counter-

³ There seems to be some uncertainty with respect to the role of constituency: does the equivalent constraint apply to words or "elements", as appears from these two examples and the formulation in Poplack (1980: 586), or to constituents, as in Sankoff & Poplack (1981: 5, cited above), and elsewhere (Nait M'barek & Sankoff, 1987: 145; Poplack & Sankoff 1988: 1175)? The predictions differ widely, if one considers that two languages may have a parallel constituent order (e.g. NPsubj-V) but different constituent internal word order (e.g. N-DET and DET-N).

examples were abundant in many other data corpora. Concerning the free morpheme constraint, agglutinative languages in particular yield numerous counter-examples showing intra-word switches (Myers-Scotton, 1993b: 31). Examples can be found in Lehtinen's 1966 study of Finnish/English and intra-word switches are also present in the Swedish/English data provided by Hasselmo in the early 1970s. On the other hand, it was considered that the Equivalence Constraint would severely restrict the switching possibilities when the two languages involved have strongly divergent word orders. This prediction was proven false when codeswitching involving typologically dissimilar languages was studied in the 1980s, e.g. Adanme/English (Nartey, 1982), Japanese/English (Nishimura, 1983), Moroccan Arabic/French (Bentahila & Davies, 1983) and Spanish/Hebrew (Berk-Seligson, 1986). In fact, in one of the first articles advocating the two constraints, it was acknowledged that "it is not quite clear how the free morpheme constraint might operate in a situation involving English and some highly inflected or agglutinative language, nor what might be the scope of the Equivalence Constraint for languages with highly different word orders" (Sankoff & Poplack, 1981: 7).

In several publications since 1988 (e.g. Poplack & Sankoff, 1988; Poplack, 1990), the Equivalence Constraint is presented as only one of a series of strategies. Apart from 'smooth switching at equivalence sites', there is 'constituent insertion', insertion of so-called 'nonce borrowings' and 'flagged switching', all of these serving to explain away many counter-examples to the Equivalence Constraint. NONCE BORROWINGS are singly occurring words from the other language that, unlike real borrowings, can be phonologically unadapted and are not necessarily widely distributed in the speech community. They are syntactically and morphologically integrated in the host language just like host language words (Poplack, Wheeler & Westwood, 1987; Poplack, Sankoff & Miller, 1988). FLAGGED SWITCHING is marked by a pause, repetition or metalinguistic comment. In Poplack, Wheeler & Westwood's (1987) Finnish/English corpus, a number of single word switches that failed to display morphosyntactic integration (and for this reason could not be classed as nonce borrowings) called for the recognition of flagged switching as a distinct category.

A major source of motivation to recognise CONSTITUENT INSERTION has been Nait M'barek & Sankoff's (1988) study of Moroccan Arabic/French. French declarative sentences have the word order subject-verb, whereas Arabic allows for both SV and VS, which leads to recurrent violation of the Equivalence Constraint, cf. (7).

- (6) *à l'époque où les Arabes* weṣl-u ḥetta *l'Andalousie*
 at DEF-time when DEF*PL-Arab arrive-PL to DEF-Andalusia
 "At the time when the Arabs reached Andalusia."
 Moroccan Arabic/French (Nait M'barek & Sankoff, 1988: 145)

- (7) *ʕemmer-hūm ža-w-ek les plats que tu fais ici*
 never-3PL come-PL-2SG DEF·PL dish REL 2SG make here
 “You never liked the dishes you make here.”
 Moroccan Arabic/French (Nait M’barek & Sankoff, 1988: 145)

While example (6) can be described as an instance of linear, left-to-right switch under equivalent word order from a French Subject NP *les Arabes* “the Arabs” to an Arabic verb *weşlu* “they arrived”, such an analysis is not possible for (7), because the word order Verb-Subject in this example is not found in French declarative sentences. But if *les plats* in this example is taken to be an inserted constituent, it no longer challenges the equivalent constraint - the relative clause *que tu fais ici* may be considered a switch under equivalent word order, since Arabic and French both have postnominal relative clauses. In addition, French nouns in Moroccan Arabic/French tend to be inserted with French definite articles (cf. Ch. 2. section 1). This often leads to a word order that seriously challenges the Equivalence Constraint. Sankoff & Poplack’s (1988) solution was to identify constituent insertion as a separate phenomenon. They did not specify exactly what kinds of constituents are liable for insertion.

The phenomenon of inserting content words and constituents was thus acknowledged, but excluded from the concept of ‘real’ codeswitching at equivalence sites. This constellation leaves little space for falsification of the Equivalence Constraint, as there is a strong general tendency for codeswitching to occur either at constituent boundaries or as content word insertion. Whenever both languages happen to permit the attested word order, as in (6) above and (8) below, such switch sites can be said to conform to the Equivalence Constraint; in case of divergent word order like in (7), the ‘same’ phenomenon will be called constituent insertion or nonce borrowing. After all, examples (7), (6) and (8) could just as well be described uniformly as instances of Subject NP insertion.

- (8) *years ago people se iban a trabajar*
 “Years ago people would go to work.”
 Spanish/English (Poplack, 1981: 176)

The applicability of the Equivalence Constraint was more or less restricted to languages with similar word order, whereas “in typologically different languages, word order incongruence makes code-switching problematic because the resulting code-switched sentence risks violating the patterns of both languages” (Poplack & Sankoff, 1988: 1177; Poplack, 1990: 55). The bilingual speakers of typologically different languages thus resort to nonce borrowing and constituent insertion.

While the introduction of insertion as a distinct possibility explained most counter-examples and thus solved the under-prediction of the Equivalence Constraint, it did not solve the problem of over-prediction. According to the constraint there is no restriction on the insertion of single function words, as long as the word order of both

languages is respected. However, with the exception of conjunctions and discourse markers, the insertion of single function words is rather uncommon. Despite these problems Poplack remained loyal to her original Equivalence Constraint, although the alternative possibility of insertion annihilated most if not all of its predictive value.

Concerning the Free Morpheme constraint, if constituent insertion and nonce borrowing are considered as distinct from codeswitching, there is scarce motivation to reject the free morpheme constraint either. After all, most bound morphemes attach themselves to either inserted words or constituents. Thus the Free Morpheme constraint is used diagnostically to distinguish between nonce borrowings, which are morphologically integrated, and single word switches, which are not integrated or 'bare', cf. Sankoff, Poplack & Vanniarajan (1991: 201). Morphological integration, however, is primarily a function of the overall productivity of the morpho-logical process at stake; I will return to this in section 4.2 below.

Poplack's methodology and terminology have influenced a whole generation of researchers on codeswitching. Much of Poplack's work is concerned with the interface between PATTERNS of bilingual language use and social variables. The investigation of patterns depends heavily on quantitative analyses of data corpora. Besides the linear approach, Poplack's work is therefore characterized by large corpus studies, careful quantification of the data and the identification of tendencies rather than strict rules.

1.3.1.2 General objections to the linear approach

The formulation of word order constraints by Pfaff, Lipski and Poplack was particularly inspired by the low frequency of occurrence of 'switching' (inserting) attributive adjectives in Spanish/English codeswitching, while there was no such restriction for predicatively used adjectives. This inhibition was attributed to the fact that adjectives frequently follow the head noun in Spanish, and normally precede it in English. Treffers-Daller (1994: 146-51) points out why difference in word order (or the Equivalence Constraint) does not explain this inhibition. Even in codeswitching between languages that share the word order Adjective-Noun, the insertion of attributive adjectives is restricted, while insertion of adjectival predicates is favoured, cf. Seaman (1972: 162) on Greek/English and Backus (1992: 54) on Turkish/Dutch, in addition to the references mentioned in Treffers-Daller (1994: 150). In other cases, the insertion of adjectives appears to be fairly unconstrained despite divergent word orders, see Androulakis (1994: 361-3) on Greek/French in Paris.

One of the primary objections to the linear, as opposed to the insertional, approach is that the former leads to a rather roundabout description of codeswitching within sentences. For instance, Sankoff & Poplack (1981: 34) observe that one of the favourite switch sites is between a determiner and a noun, while in fact single nouns are frequently inserted irrespective of the preceding word. It is more economical to speak of noun insertion than to enumerate all word categories that may occur adjacent to a noun. Likewise, if one were to investigate codeswitching in grammatical relations such as between a verb and its Object complement (that is, Object NP insertion) in

a strictly linear approach, one would overlook all the cases where another constituent, in either language, intervenes between verb and Object NP. To simply disregard the “intervening material between the switch and the adjacent syntactic categories”, as Poplack (1980: 597) proposes to do with movable constituents like adverbs and adverbials, is really an artificial solution. Moreover, the intervening constituent may not be movable (e.g. an indirect Object NP).

Another difficulty is that the linear branch in codeswitching studies, as exemplified by Poplack’s work, is in reality often an insertional approach in disguise. This has become apparent with the introduction of inserted nonce borrowings and constituents in Poplack and Sankoff’s publications. In fact the element of insertion, and consequently, that of matrix language, was implicitly present from the beginning. For example, all the ‘intrasentential’ switches in Poplack’s (1980: 602) overview table of her quantitative data analysis must be interpreted as insertions. To cite another example, Safi (1992: 75) apparently means insertions when she talks about ‘switched items’. Yet the existence of a matrix language is not made explicit or, still further, it is even explicitly denied (Poplack, 1980: 597, n.13).⁴ Of course there can be no insertion without an embedding structure. In fact the term ‘intra-sentential codeswitching’ suggests a sentence structure, however the adherents of the linear approach are reluctant to assign this structure to either language.

Apart from this apparent contradiction, there is the terminological inconsistency. Insertion and juxtaposition of elements from two languages are not always clearly distinguished, and the term ‘switch’ is applied to both. Expressions such as “to switch from (language A) to (language B)” may denote a transition in the linear, left-to-right sense, or the insertion of a language B element into a language A frame. The latter meaning is usually associated with single words (as in Nortier, 1990), and the former with switching between clauses or sentences. This leaves considerable ambiguity in the case of constituents: are switch sites being referred to, or inserted constituents?⁵

The objection to a linear description of codeswitching is not merely a matter of aesthetic preference: a lot of confusion is generated by the absence of a distinction between matrix and embedded language. To illustrate this I will again refer to the question of adjective placement in Spanish/English. Pfaff (1979) states that “adjective/noun mixes must match the surface word order of both the language of the adjective

⁴ In the same passage (1980: 597), Poplack talks about a *base* language but its definition is unclear.

⁵ A careful definition of switch is extremely important when codeswitching patterns are to be quantified. In principle, one insertion is worth two switch sites, although different interpretations are possible if the inserted constituent happens to be the first or last in a clause, sentence, utterance or discourse. Since there is no uniform solution, and the choices made are not always clarified, one has to be very cautious with interpreting and comparing the quantitative analyses available in the literature.

and the language of the head noun”. She then proceeds to say that the example in (9) from her own data is not a valid counter-example because *toast* is actually a loanword, and to support her claim she cites the constructed example (10), which had been rated unacceptable in a gram-maticality judgment test (Pfaff, 1979: 306-7).

- (9) me huele a *toast* quemado
“It smells of burned toast to me.”
- (10) **I went to the house* chiquita.
“I went to the small house.”

But an alternative explanation of why an English noun followed by a Spanish adjective is acceptable in (9) but not in (10) is that in (10) the matrix language is English and therefore English word order is required. Besides, there seem to be restrictions on the insertion of attributive adjectives generally, as opposed to nouns. Another example of the arising due to both the neglect of the concept of matrix language and the terminological inconsistency just mentioned is the following: Choi (1991: 889) cites Joshi who states that “closed-class items cannot be switched” (Joshi, 1985: 194, cited on p. 35 hereafter). Choi took Joshi’s statement as a constraint on the juxtaposition of a closed class item from one language and any word from another language. Choi presents (11) and similar examples of Korean nouns in otherwise English contexts to challenge Joshi’s claim, as they contain a switch between a determiner and a noun. But Joshi, who can be considered to be one of the founding fathers of the current matrix language approach (see 3.3 below), would probably find (11) in accordance with his predictions, because he would identify English as the matrix language, and consider *nokum* a ‘switched’ open-class item. In Joshi’s terms, for an item to be ‘switched’ means to be in another language than the matrix language (i.e., to be inserted, in the terminology that I will employ). Exactly the same misunderstanding of Joshi’s paper is found in Belazi, Rubin & Toribio (1994: 227).

- (11) *I command you to do the* nokum!
“I command you to do the recording!” Korean/English (Choi, 1991: 889)

The linear approach makes no principled distinction between content and function words. In principle, crosslinguistic congruence of word and constituent categories is crucial for the identification of possible switch points under the Equivalence Constraint. For instance, the suggestion that switching is possible between a noun and an attributive adjective if both languages have the word order noun-adjective presupposes that the ‘adjective’ and ‘noun’ categories are the same for both languages. Congruence has only become an explicit concept in the more recent work of Poplack and her associates, where it is referred to as STRUCTURAL EQUIVALENCE (e.g. Sankoff, Poplack & Vanniarajan, 1991; Meechan & Poplack, 1994; Poplack &

Meechan, 1995). However, in accordance with the Equivalence Constraint, structural equivalence is viewed primarily as a matter of word order.

1.3.2 Government approaches

While the linear approach of Poplack and her associates assumes that codeswitching takes place in ‘surface structure’ (Sankoff & Poplack, 1981: 12), others have tried to formulate rules in terms of structural hierarchy, or dependency between words. Woolford’s study (1984) was basically a reformulation of the Equivalence Constraint in Government and Binding terminology.

1.3.2.1 Di Sciullo, Muysken & Singh’s Government Constraint

The most pronounced claim was made by Di Sciullo, Muysken & Singh (1986), and formalised as the GOVERNMENT CONSTRAINT:

(12) if X governs Y, ...X_q ...Y_q ... (Di Sciullo et al., 1986: 5)

In other words, “whenever constituent X governs Y, both constituents must be drawn from the same language” (Muysken, 1987: 365). Typical cases of government would be case assignment or subcategorization. This government constraint was an easy target for falsification. Complements of verbs and appositions are in fact amongst the most common switch types. In many instances the government constraint is saved by the assumption of a LANGUAGE CARRIER: When a language q (L_q) governor, such as a transitive verb, governs a language p (L_p) complement (noun) and the latter is accompanied by an L_q marker, Di Sciullo, Muysken & Singh consider the government constraint to be saved. It seems that in principle any function morpheme (determiner, case marker, plural marker) can function as an L_q marker (Muysken, 1991: 269), so that (13) becomes possible under the government constraint. This solution does not work, however, if the governed constituent consists of a single content word, or if the governed constituent is entirely in the L_p as in (14).

(13) ha ricevuto il *diplôme*
 have·PRES·3SG receive·PART DEF·M diploma
 “She has received the diploma.”
 Italian/French (Di Sciullo, Muysken & Singh, 1986: 13)

(14) wara-l-u le *bulletin*
 show·PERF-to-3M DEF·M school·report
 “He showed him the school report.”
 Algerian Arabic/French (Keddad, 1986: 242)

1.3.2.2 Belazi, Rubin & Toribio’s Functional Head Constraint

A rather different conception of government is maintained by Belazi, Rubin & Toribio (1994), who base their claims on spontaneous data and grammaticality judgments on constructed examples in Spanish/English and Tunisian Arabic/French. In their approach, functional categories such as Complementizer, Inflection, Negation and Determiner are considered heads of phrases and codeswitching between these and their complements is inhibited. This is formalized in their FUNCTIONAL HEAD CONSTRAINT: “The language feature of the complement f-selected by a functional head, like all other relevant features, must match the corresponding feature of that functional head” (1994: 228). As testable examples Belazi et al. claim that “switching is disallowed between a C⁰ and its IP complement (...) and between D⁰ and NP” (1994: 229). Hence the ungrammaticality of their example (14), repeated here as (15). Belazi et al. admit that a case like (15) would naturally be acceptable as an instance of borrowing.

- (15) **He is a demonio*
 “He is a devil.” English/Spanish (constructed) (Belazi et al., 1994: 227)

- (16) (*c’est le fer qui donne ..*) lli yi-ħa:rib *l’anémie*
 it-is DEF-M iron REL gives .. REL 3M-fight DEF-anemia
 “It’s iron that gives, .. that fights anaemia.”
 Tunisian Arabic/French (Belazi et al., 1994: 226)

Belazi et al. might have been encouraged by the fact that in Tunisian Arabic, French nouns tend to be inserted together with the French definite article like *l’anémie* in (16), just like in Moroccan and Algerian Arabic (cf. examples (7), (6) and (14) above and the discussion in Ch. 2 section 1.1). Likewise, the Functional Head Constraint would be valid for certain complementizers in certain languages, but it does not represent an overall tendency. Recurrent counter-examples such as switches between determiners and nouns, or between complementizers and complement clauses will be found in many data corpora (see, for instance, examples (11) and (13) above and (17), (32) and (39) below), whereas nominal and verbal inflection from one language can in general combine with lexical heads from another language; inflectional affixes from agglutinative languages, in particular, tend to be productive with foreign stems (see (18), (20), (22), (24), (31) and (43) below). Belazi et al.’s claim is diametrically opposed to the matrix language approaches, which basically predict that function morphemes originate from one language and content words from another, as I will explain shortly.⁶

As it happens, both governed and ungoverned, as well as governing content words and constituents can be ‘switched’ (inserted), so that, while it will be possible to

⁶ Cf. also the criticism in Mahootian & Santorini (1996).

describe codeswitching in GB terminology, the government principle itself does not seem to constrain codeswitching.

1.3.2.3 Government approaches that include insertion

More successful, in terms of making the right predictions, are the proposals in which heads of phrases control morpheme order and case marking, instead of prohibiting codeswitching altogether. Such proposals have been developed independently by Pandit (1990), Santorini & Mahootian (1995; Mahootian & Santorini, 1996), Halmari (1993) and Muysken (1995). In Pandit (1990: 43) it is stated that “code switching must not violate the grammar of the head of the maximal projection within which it takes place”, or, as Santorini & Mahootian (1995: 9) put it “the language of the head determines the phrase structure position of its complements in codeswitching just as in monolingual contexts”. Similarly, Halmari (1993: 1061) on Finnish/English: “An English lexical item can be inserted in the terminal node, provided that when a government relation is involved, Finnish morphosyntactic rules are not violated.” Whether these statements make the right predictions with regard to codeswitching naturally depends on which elements are identified as phrasal heads. The relative order of noun and determiner, for example, is predicted more reliably if the determiner (Det) is identified as the head, as in Santorini & Mahootian (1995), rather than the noun, as in Pandit (1990: 39). In (17) we see the Swedish postnominal definite article attached to an English noun.

- (17) betal-ar du skatt-en å *insurance* på *building-s-en* (..)
 pay-PRES 2SG tax-DEF and insurance on building-PL-DEF (..)
 “You pay the tax and insurance on the buildings (..)”
 Swedish/English (Hasselmo, 1970: 198)

On the other hand, Pandit (1990) counts verbal inflection (INFL) as a phrasal head. This would help to account for switches such as the Swahili/English *wa-na-vyo-behave* in (18). However, Pandit asserts that INFL and the verb must be from the same language, as “the attachment of INFL from one language to the verbal complex of another language would violate the phonetic or morphological system of the language of the head” (1990: 61 n.12). As for Santorini & Mahootian, they “do not consider inflectional features like [tense] as separate heads in phrase structure, but assume that they are instantiated as syntactic features on lexical heads” (1995: 5). Example (18) is ruled in by Halmari, who allows for lexical insertion and identifies the Agreement node as an independent governor in addition to verbs and adpositions (1993: 1062). Consequently, the verbal inflection projected by AGR has to be in the same language as the non-lexical elements (in casu: determiners and case marking) in the Subject NP.

- (18) u-na-anza ku-behave kama watu wa huko
 2SG-NONPAST-begin INF-behave as people of there

wa-na-vyo-behave

3PL-NONPAST-MANNER-behave

“You begin to behave as people from there behave.”

Swahili/English (Myers-Scotton, 1993b: 103)

In his 1995 article, Muysken reviews the linear, government, and matrix language approaches, and attempts to reconcile the major tendencies in the grammatical analysis of codeswitching. He presents an amendment to the 1986 Government Constraint:

* [Xp, Yq], where *XL*-marks *Y*, *p* and *q* are language indices, and there is no equivalence between the category *Y* in one language and the category *Y* in the other language involved (Muysken, 1995: 195)

L-marking is “a more restricted notion of government by a non-function word under thematic marking” (Muysken, 1995: 187), i.e., L-marking applies to the complements of a V, N, A or P (the symbol *L* derives from ‘lexical category’). As the author recognises (1995: 187), L-marking, like the original Government Constraint, is too strong in that it excludes the insertion of verbal and prepositional complements (cf. ex. (14) above and (19) hereafter), so now complements can be inserted under the condition of equivalence.

- (19) žib-li-ya een glas water of zo
 bring-IMP-to-1SG INDEF glas water or so
 “Get me a glass of water, or something.”
 Moroccan Arabic/Dutch (Nortier, 1990: 131)

Muysken leaves unspecified whether he refers to structural equivalence or word order equivalence à la Poplack, which he considers to be a subcategory of structural equivalence.

The replacement of the original Government Constraint by L-marking eliminates some of the problems that arose from the identification of certain functional elements as governors. In particular, the original constraint inappropriately ruled out switches between INFL and the Subject and between Det or Q[uantifier] and N. Note, however, that the equivalence factor could just as well have been added to the original constraint to repair its underprediction. Because L-marking affects constituents, the new ‘L-marking Constraint’ has little to say about content word insertion.

One drawback affects all of these ‘insertion variants’ of the government model: while they generally allow for the insertion of content words, difficulties arise when the inserted word itself happens to be a governor (or L-marker). Inserted verbs are

common examples - assuming that it is the inserted lexical element that governs its complements, rather than the verbal inflection from the matrix language. In order to explain (20)-(24) within these government frameworks we must assume that the complements of the embedded verbs are inserted under equivalence. Even if we are prepared to assume crosslinguistic equivalences for all kind of verbal complements it appears that word order is in accordance with the language of the verbal inflection rather than the language of the lexical verb, see for instance (23) where we find the Hindi word order Complement-Verb.

- (20) *me pretend-attiin olevan elefantteja*
 we pretend-PAST·1PL to·be elephant·PL·PARTITIVE
 “We pretend to be elephants.” Finnish/English (Halmari, 1993: 1049)
- (21) *ngóh take sei fo*
 I take subject four
 “I take four subjects” Cantonese/English (Gibbons, 1987: 58)
- (22) *feedáil sí é*
 feed she him
 “She fed him.” Irish/English (Stenson, 1990: 195)
- (23) *parents te depend hona ě*
 parents on depend be/become AUX
 “It depends on the parents.” Hindi/English (Romaine, 1995: 135)
- (24) *ro(h)o yake i-ko-na-mu-nuire eh?*
 greediness his it-COP-TMA-him-hurt TAG
 “(It is clear that) his mind troubles him.”
 Shaba Swahili/French (De Rooij, 1996: 106)

To summarize the merits and demerits of the government approaches: firstly, as long as functional categories are identified as heads of phrases, and content word insertion as well as constituent insertion is allowed for, codeswitching data corroborate the government model. Secondly, government models make no predictions concerning the possibility of switching (inserting) constituents that are neither governors nor governed. Thirdly, in the case of inserted governing verbs, government models frequently make the wrong predictions.

1.3.3 Insertion models

Neither the linear word order constraints, such as Timm’s (1975) five constraints and Poplack’s Equivalence Constraint, nor the subcategorization or government-based constraints make a principled distinction between the roles of the two languages

involved in codeswitching. As for content and function words, their difference in status was noted as early as 1966 by Lehtinen (1966: 175-7), but often neglected in later studies. Crosslinguistic congruence of word and constituent categories is finally receiving increasing attention in recent years. The asymmetric roles of the two languages, the differentiation of content and function words and the aspect of categorial congruence (or some of these three elements) are implicit in many studies. In the following I will argue that they are best embodied in the insertion approach to codeswitching. I have chosen to discuss in detail the work of three scholars, Hasselmo, Bautista and Myers-Scotton, as they offer a more comprehensive treatment of codeswitching. Hasselmo, who worked on American Swedish in the late sixties and the seventies, did not claim any validity for his model beyond the bilingual Swedish immigrant communities he studied. His model is attractive for its elegant format and the element of implicational hierarchy. Bautista was very early in formulating an insertion model for Tagalog insertions in English, and English insertions in Tagalog. Of particular interest is her observation that the insertion of constituents is subject to constraints of structural equivalence, which runs contrary to the commonly held belief that the internal make-up of embedded constituents is independent of the matrix language. In the last decade Myers-Scotton's work has been prominent in formulating the matrix language perspective. Besides Hasselmo's, Bautista's and Myers-Scotton's publications I will discuss Joshi's and Schmid's contributions to the discussion on matrix languages. Joshi's papers were an important source of inspiration for Myers-Scotton's work and Schmid emphasises the selectional properties of the matrix language more than the other scholars do.

1.3.3.1 Hasselmo's Ordered Selection

In a number of publications (1972a, 1972b, 1974, 1975) Hasselmo presented a model for bilingual speech behaviour among American Swedes which he called ORDERED SELECTION. This model describes how English elements can be integrated in Swedish utterances.

Starting from a lexical core a constituent is built in subsequent stages or levels. At each level a choice is made for either English or Swedish. In a noun phrase for instance, the noun itself, derivation, inflection for plural and marking for definiteness are subsequent levels. The 1974 and 1975 versions of the model added prosody as the final level. Thus an English lexeme (first level) may receive either English or Swedish derivational suffixes at the second level. As an example Hasselmo (1974: 150) provides both *saloon-keep-er* and *saloon-keep-are*, the latter containing the Swedish agentive suffix. However, once a choice for Swedish is made, all subsequent levels must be Swedish (the INTER-LEVEL RESTRICTION). Firstly, this means that a Swedish lexeme only receives Swedish derivation. Secondly, Swedish derivation can only be followed by Swedish inflection (level 3), so that *saloon-keep-are* has the Swedish plural *saloon-keep-are-ø* (which happens to be unmarked in this case) and the English suffix is excluded: **saloon-keep-are-s*. In other words, only English

elements embedded in Swedish morphological and syntactical structures are permitted, and not vice versa. Hasselmo's approach is a kind of matrix language model in which Swedish is consistently assumed to be the matrix language.

In addition Hasselmo claims that within the same level all elements must be from one language (the INTRA-LEVEL RESTRICTION). The result is an implicational scale: if English is selected for a given level, all preceding levels within the same constituent must be English.

Stages and selections	'Pure' Swedish	Intermediate Combinations				'Pure' English
Lexemes	S	E	E	E	E	E
Derivation	S	S	E	E	E	E
Word inflection	S	S	S	E	E	E
Function words	S	S	S	S	E	E
Prosody	S	S	S	S	S	E

Table 1.1. Ordered Selection
(adapted from Hasselmo, 1974: 144; 1975: 252)

This is a structural model: "Ordered selection is assumed to operate independently of the sequential order of elements" (Hasselmo, 1975: 259).

Ordered Selection is also applied to the major sentence constituents. For instance, the verbal inflection can be English only if the entire Object noun phrase is English (and the verb stem is English), that is to say, if the Object NP is English, the entire VP can be English too. Likewise, the Subject NP can only be entirely English, i.e. contain for instance the English definite article, if the entire VP is English (1972b: 270). Note that the implicational hierarchy in the domain of the sentence constituents is independent from that in the domain of the constituent: an English lexeme can occur in Subject position as part of a Swedish NP even though the VP is entirely Swedish. That is, an English lexeme can still be selected on the first level of constituent assembly even if this particular constituent represents the last level in the sentential hierarchy, and the previous level constituents are all realised in Swedish.

One problem concerns the intra-level restriction which states that within a particular level all elements must be from the same language. This premises is primarily based on the 'ungrammaticality' of forms like *fox-y-het* that contain both an English derivational suffix (-y) and a Swedish one (-het), as opposed to *fox-ig-het* and the all-English *fox-y-ness* (Hasselmo, 1972b: 265). But if any Swedish

derivational affix is at all productive with English words, why would it not be productive with English words that contain some (English) derivational affix I suspect that Hasselmo found no counter-examples because, on the whole, Swedish derivation is barely productive with English words. Likewise at the inflectional level: since neither Swedish nor English verbs, nouns and adjectives receive more than one inflectional suffix anyway - Hasselmo categorises the Swedish definite suffixes as function words -, the intra-level restriction cannot fail at this level. That the restriction does go wrong at the first level of selecting lexemes is illustrated by example (17) above. In this fragment, taken from one of Hasselmo's earlier publications, a Swedish and an English noun occur in coordination.⁷

Concerning the inter-level restriction, everything depends on the appropriate distinction of levels. Hasselmo designed his model in the particular context of American Swedish. However, a wider applicability of the ordered selection principle can be achieved by allowing for more than one level of derivation, inflection and function words. Several versions of the model actually differentiate between word inflection and phrase inflection (Hasselmo, 1972a: 170), or recognise successive levels of function words (ibid.; 1974: 150) to account for the 'grammaticality' of such examples as *i the colleges* where the Swedish pre-position *i* "in" and the English definite article are both function words. As a consequence the intra-level restriction becomes superfluous, since one can usually assume successive levels of derivation, inflection, etcetera.

The inter-level restriction does not allow for insertions within insertions. This restriction is adequate if one language consistently serves as the matrix language in which elements from the other language are inserted. This may very well have been the case in Hasselmo's American Swedish data, as it is a common type of code-switching. Yet there are varieties of codeswitching where both languages function at different times as matrix language, and in such a context we also encounter successive layers of insertion. In (25) there is a first level-to-level switch between the French NP and the Arabic function word *f*, and at the subsequent level a switch back from Arabic to French is needed to embed the PP *l'armée de l'air* in the French clause (see also ex. (41) and (42) and their discussion at the end of this chapter).

- (25) *je devais faire pilote f l'armée de l'air*
 I would do pilot in DEF-Air Force
 "I was going to become a pilot in the Air Force"
 Moroccan Arabic/French (Lahlou, 1991: 254)

At the clause or sentence level the Ordered Selection model is on the whole less tenable. If constituents are not all in the same language, a hierarchy among them must

⁷ I assume that the intra-level restriction applies to coordinated elements; otherwise the restriction has no meaning at all for the level of lexeme selection.

be assumed, with the English constituent always representing the first level of language selection. The same hierarchical ordering cannot be consistently applied to different Swedish sentences containing different English constituents. Besides, the inter-level restriction rules out the insertion of two constituents in the same sentence, unless they are subsequent in the hierarchical ordering of levels. The INTRA-level restriction excludes the possibility of coordinated constituents in different languages. I suspect that Hasselmo did not intend to pose these ‘constraints’. Even if they are not challenged by his American Swedish data, they are certainly not universally applicable. The following example shows co-ordinated NPs in French and Dutch:

- (26) *il y avait des colporteuses qui étaient là, avec [pause] ne selder,*
 EXIST·PAST INDEF·PL saleswoman REL were there with a celery

ne porei, een ajuintje of twee et deux trois carottes
 a leek a onion or two and two three carrots

“There were door-to-door saleswomen who were there, with a celery, a leek, one or two onions and two, three carrots.”

Brussels Dutch/French (Treffers-Daller, 1994: 206)

Hasselmo presumably posited a hierarchy of constituents to account for the absence of singly occurring English inflected verbs as opposed to nouns and the absence of English Subject noun phrases as opposed to Object noun phrases in otherwise Swedish sentences. But these two phenomena may have different explanations. The restriction on inserting English inflected verbs can be due to a general constraint on the insertion of inflected word forms, to which the insertion of plural nouns is a common exception. Inserted Subject NPs have been attested in other data corpora (e.g. Poplack, 1980: 602 for Spanish/English, Treffers-Daller, 1994: 207 on Dutch/French, and examples (7) and (6) above) and the fact that Objects are inserted more often may have other explanations. Hasselmo himself suggests that this may have to do with a tendency to locate new information in the Object position (1975: 259). This explanation presupposes that there is a relationship between constituent insertion and coding new information. I will return to this matter in the final chapter.

Unfortunately Hasselmo did not provide samples from his recordings of American Swedish to illustrate and confirm the implications of his model. Instead, he used acceptability judgements. As he noted himself, some patterns implicated by the model are not supported by examples from the data, although serious counter-examples (selection of English after a previous selection of Swedish) appear to be few in number (1972b: 270).

Despite these critiques Hasselmo must be credited with providing a plain and elegant model that mostly makes the right predictions. His model provides a uniform principle for the insertion of single content words, inflected content words, inflected content words accompanied by some function words and, finally, complete

constituents. The implicational hierarchy rightly prohibits the insertion of single function morphemes. At the same time it mistakenly excludes the possibility of successive layers of insertion. Underlying the model there seems to be the idea that constituent assembly takes place ‘bottom-up’ from a lexical core through a number of subsequent stages. Yet nowhere does Hasselmo explicitly make this claim: his levels refer solely to the implicational order. The Ordered Selection model clearly distinguishes the roles of the two languages and assumes a different status for content words and other morphemes. Crosslinguistic congruence of categories is not mentioned.

Hasselmo was ahead of his time and maybe this is one of the reasons why his model has attracted relatively little attention. An exception is a paper by Petersen (1988) on a Danish/English bilingual child. This paper bears clear traces of influence from Hasselmo’s work. Discussing only word-internal switching she poses the following hypothesis:

The dominant language hypothesis states that in word-internal code switching, grammatical morphemes of the dominant language may co-occur with lexical morphemes of either the dominant or the non-dominant language. However, grammatical morphemes of the nondominant language may co-occur only with lexical morphemes of the nondominant language (Petersen, 1988: 486)

In Petersen’s hypothesis the term ‘dominant language’ refers to the language the speaker knows best, which is Danish in the case of the bilingual child studied. Hasselmo simply observed that English lexical morphemes could be combined with Swedish grammatical morphemes and not vice versa. Petersen made the same observation, but related the asymmetric roles of the two languages to the concept of language dominance, thus adding an important generalisation. Note that Petersen’s formulation is a step forward compared to Poplack’s free morpheme constraint. Hasselmo’s American Swedish and Petersen’s concept of a dominant language have much in common with the concept of a matrix language (ML) in the matrix language approaches discussed hereafter. As will become clear in the course of the study of Moroccan Arabic/Dutch CS in Part II, however, there is no tendency for the ML to be the most fluent language of the speaker; this point will be discussed in Chapter 11.

In connection with Petersen’s hypothesis, it should be further noted that the number of productive morphological processes that can be involved in ‘word-internal code switching’ differs from one language to another (see 4.2 below). For the comparison of Danish and English, for example, it is relevant that Danish has nominal suffixes as definite articles, whereas English only has free form articles. Therefore an English noun occurring in a Danish NP is likely to confirm Petersen’s hypothesis while a Danish noun in an English NP will never challenge it.

1.3.3.2 Bautista's model of bilingual competence

In her 1975 dissertation Bautista designed a model of the bilingual speaker. She used a corpus of radio broadcasts in which codeswitching between Tagalog and English was commonplace. As far as I know, this dissertation, which was published in 1980, and the abbreviated version published in 1975 have received little or no attention outside the Philippines. The reason for including a discussion of Bautista's model here is that, while it is very similar to the much more influential matrix language approaches to be discussed in the next subsection, it has some attractive features not shared by all other proposals. Firstly, unlike Hasselmo's Ordered Selection and the most influential matrix language approach, Myers-Scotton's Matrix Language Frame model, Bautista's model assumes that in principle either of two languages can provide the syntactic frame and conversely, material of either language can be inserted. In other words, Bautista does not a priori designate one of the bilingual speaker's languages as the matrix language. Secondly, Bautista posits congruence requirements for the insertion of constituents, thereby recognising that constituent insertion does not occur at random.

Finally, the conception of her competence model is quite akin to the psycholinguistic theories of speech production that underlie the later matrix language approaches, including the present study. The central feature of both Bautista's model of the bilingual speaker and the psycholinguistic model of speech production in Levelt (1989), among others, is modularity: several aspects of the speech production process are handled by autonomous components. The relative autonomy of each component explains the fact that elements from two languages may be combined into a constituent according to the rules of one language. Thus an English noun, for instance, may combine with Tagalog function morphemes to form a Tagalog nominal constituent, and English and Tagalog constituents may be combined into an English or Tagalog sentence.

Bautista's bilingual competence model is formulated in terms of Chomsky's 1965 version of Transformational Grammar, also known as the Standard Theory. The bilingual model is an amendment of the monolingual model. It can be outlined as follows (see *Fig. 1.1*): A 'primitive S' sentence is developed by Phrase Structure rules (contained in the Phrase Structure Component) and the Lexicon to produce a deep structure. This deep structure is fed into the Transformational Component, where transformations are applied cyclically until a surface structure is arrived at. The surface structure is then fed into the Phonological Component that produces a phonetic representation (Bautista, 1975: 75). (The Semantic Component, the working of which is not specified, need not detain us here.) Now the bilingual speaker simply has two Phrase Structure Components, two Lexicons, two Transformational and two Phonological Components, since each of these components is language specific. For ease of reference one set of components is labelled L1 and the other set L2, although

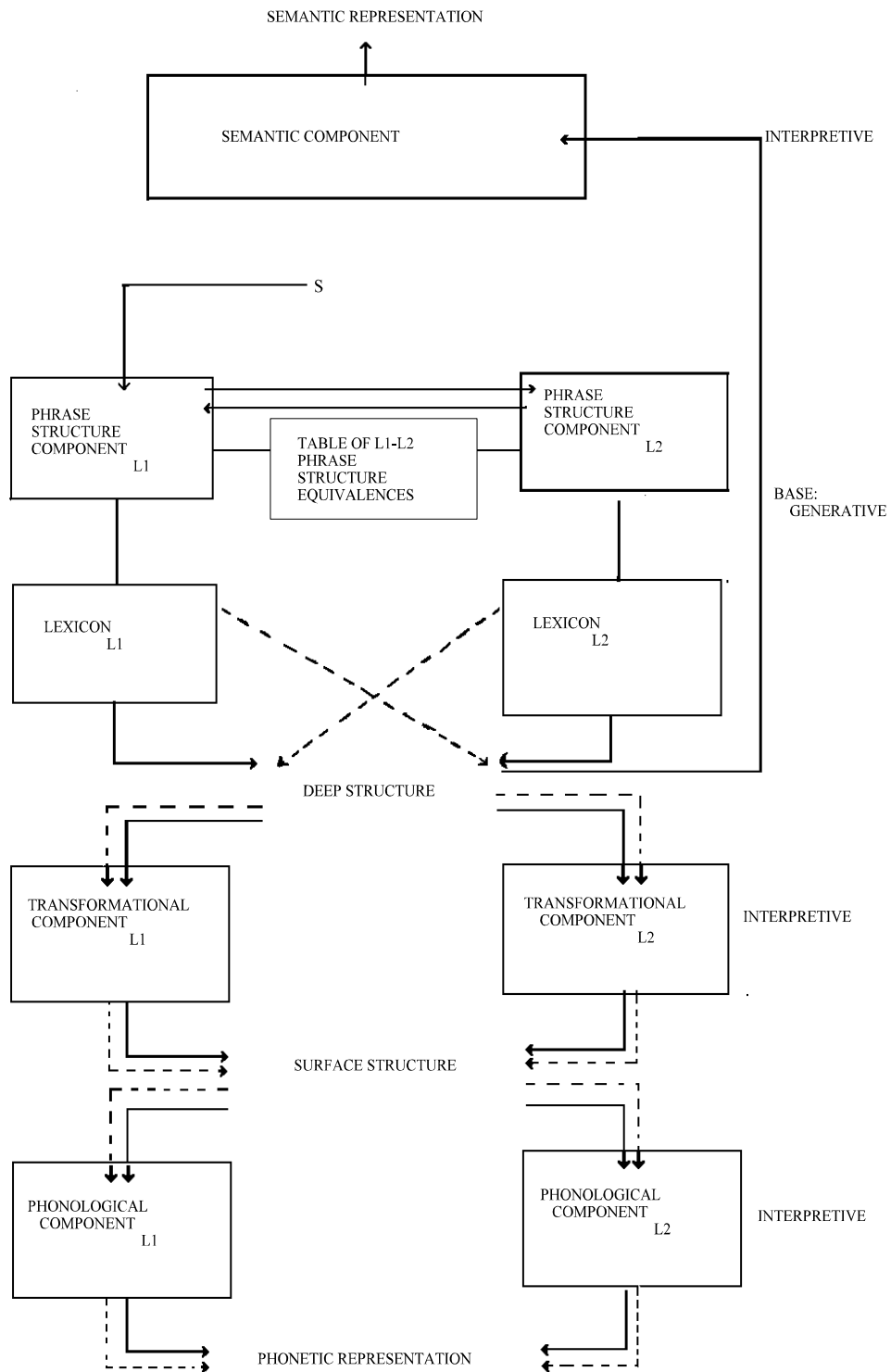


Fig. 1.1 A Model of Bilingual Competence (Bautista, 1975: 73; 1980: 210)

Bautista does not formally assume different roles for Tagalog and English.

Two types of crossover between the L1 and the L2 half of the model are identified in order to account for codeswitching. Firstly, there can be a branching into the other Lexicon, e.g. a branching into the L2 Lexicon (at some point) during the course of generating an L1 deep structure. This leads to insertion of an L2 lexical item into an L1 utterance. The inserted element will presumably undergo further treatment in the L1 Transformational Component but note that it must find its way back to the L2 Phonological Component for its phonetic interpretation. (Bautista does not address this problem.)

Secondly, there is a direct link between the two Phrase Structure (PS) Components. The L1 component can ‘delegate’ part of its work to the L2 component and vice versa. For instance when PS Component_{L1} rewrites an initial S as [S Subordinate-Conjunction S], the second S can be shunted to PS Component_{L2}. PS Component_{L2} then further rewrites this S and together with the L2 Lexicon produces an L2 deep structure. This mechanism is invoked to account for the embedding of conjoined and subordinate clauses, and for embedded phrases that, according to the Standard Theory, are derived from deep structure sentences. The transformational machinery is assumed to tidy up the surface forms consisting of embedding and embedded sentences - this part of the model is not analysed in much detail. Bautista treats the embedding of noun phrases and prepositional phrases that are not derived from sentential units in an analogous way: PS Component_{L1} delegates the development of a phrase to PS Component_{L2} (or vice versa).

Guided by her data, Bautista recognises that there is no a priori correspondence between Tagalog and English phrases. Therefore her model is equipped with an additional sub-apparatus: the TABLE OF L1-L2 PHRASE STRUCTURE EQUIVALENCES. This table includes at least the following information:

Equivalence 1	Tagalog <i>ang</i> -NP	= English NP	see (27)
Equivalence 2	Tagalog <i>ng</i> -NP-genitive	= English PP	see (28)
Equivalence 3	Tagalog <i>sa</i> -NP	= English PP	see (29)
Equivalence 4	Tagalog PP	= English PP	see (30)

Table 1.2. L1-L2 Phrase Structure Equivalences
(Bautista, 1975: 83; 1980: 218)

The last column of *Table 1.2* refers to the example sentences, which are reproduced from Bautista (1975:83; 1980:218). Unfortunately the author does not provide complete glosses; *po* in (28) marks referens towards the addressee.

- (27) *ang intensyon talaga is to maintain it in the original state*
the intention

- (28) *this is in connection with the celebration* po ng anibersaryo ng D.A.R.
REF of the anniversary of the D.A.R.
- (29) *palagay ko ay we would put in more interest* sa ating pagtatanghal
into our presentation
- (30) *(..) and some schools are conducting review classes* para sa kanila
for them

To appreciate Bautista's Table of Equivalences one needs to have some understanding of the Tagalog phrase markers *ang*, *ng* and *sa* (analogous tripartite paradigms exist for pronouns, demonstratives and proper names). The following main facts are extracted from Schachter & Otnes' (1983) grammar of Tagalog and the discussion in Andrews (1985).

Verbs take three types of arguments, which Schachter & Otnes call Actor, Object and Directional. The first two are marked by *ng*, the latter by *sa*, unless they are Topic. The Topic NP is marked by *ang*. Depending on the verb and the verbal form, the Topic argument may take any of the semantic roles Actor, Object, or Directional. Different grammatical analyses identify either the pragmatic function of Topic or the semantic role Actor with the grammatical function of Subject (cf. Andrews, 1985). In Bautista (1975, 1980) the Topic NP is called Subject.

What Bautista's first Equivalence predicts is that a Tagalog NP that occupies an NP slot in an English sentential frame is invariably in the topical '*ang*-form' (marked by *ang* for NPs other than pronouns, demonstratives and proper names), irrespective of its semantic role and its grammatical function as Subject or Complement in the English sentence. Note that this Equivalence is inferred from her data (confirmed by Sobolewski, 1982: 55) and does not derive in any direct way from Tagalog or English grammar.

As Bautista acknowledges herself (1975: 83), the information in the table should be worked out and stated more rigorously. For instance, one might inquire as to whether the English PPs in the Equivalences 2-4 can be specified and categorised in any systematic way. Also, contrary to Bautista's claim, the table works only for Tagalog phrases embedded in English sentences and not the other way round. In its present form the table suggests that any English PP can be inserted as a Tagalog '*ng*-NP-genitive', '*sa*-NP', or 'PP', a claim which is probably unwarranted. Furthermore, full English NPs that function as Topic or Complement in Tagalog sentences are rare, but there are indications that these tend to receive the Tagalog *ang* and *ng* markers (Bautista, 1980: 50, 54-6; Sobolewski, 1982: 55), which means that they do not occupy the position of an *ang*-NP in Tagalog as the table suggests. They are, at most, part of an *ang*-NP.

If we abstract from her theoretical framework, we see that Bautista describes code-switching as the insertion of lexical items, constituents, and clauses from one language into a frame set by another language. As for the insertion of constituents

(phrases), she posits a structural equivalence requirement to account for the attested restrictions and regularities. No such requirement is formulated for the insertion of lexical items. Though Bautista's (1980) detailed description of Tagalog/English code-switching reveals strikingly different roles for both languages, her Bilingual Competence Model treats the two languages as essentially equal. Likewise, 'open-set' and 'closed-set' lexical items are distinguished in the description of the data, but not in the theoretical approach.

1.3.3.3 Matrix language approaches

The idea of one language regulating most of the sentence structure evolved independently of Hasselmo's and Bautista's proposals during the 1980s. The MATRIX LANGUAGE organizes and expresses the grammatical relations in the sentence by means of inflection, function words and word order. For this reason, the distinction, roughly speaking between, CONTENT WORDS on the one hand and INFLECTION and FUNCTION WORDS on the other is an indispensable ingredient of the matrix and embedded language approach.

Several researchers have come up with various types of distinction between the languages involved in codeswitching, mentioning a base or matrix language. Early proposals include Wentz (1977), Sridhar & Sridhar (1980), Joshi (1981, 1982, 1985), Pandit (1986), and Petersen (1988). Joshi's 1981 paper was an important source of inspiration for later work, in particular the papers by Doron (1983), Ewing (1984), Klavans (1985), Schmid (1986) and Nishimura (1986), as well as the model developed in the 1990s by Myers-Scotton (to be discussed hereafter).

Sridhar & Sridhar (1980) do not define their terms 'host language' and 'guest language'. Their most pronounced claim concerns constituent insertion and is formulated in the Dual Structure Principle. From this formulation it follows that the host language can be identified by examining the constituent order in the sentence.

Dual Structure Principle. The internal structure of the guest constituent need not conform to the constituent structure rules of the host language, so long as its placement in the host language obeys the rules of the host language. (Sridhar & Sridhar, 1980: 412)

Joshi (1982, 1985) and Petersen (1988, discussed above) actually relate inflection and function words to the matrix language; Nishimura (1986) uses word order to identify the matrix language on sentence level. Pandit in her earlier work (1986) was also mainly concerned with word order, but defined her matrix language, which she called 'governing' language, as the language of the verb in the sentence (1986: 36, 98-9). According to Wentz (1977: 182, cited in Poplack 1980: 617 n.13) the 'language of the sentence' is the one in which the determiner and the main verb are produced. Klavans (1985) argues for a definition of the matrix language in terms of the inflection-bearing element of the verb. Schmid (1986) contends that the matrix language can be inferred from the applicability of grammatical rules generally (see

below). Linking up with Joshi and Klavans, Kamwangamalu (1989) presents his Matrix Code Principle. He does not offer a definition of the matrix language, but contributes a link to the sociolinguistic context: he notes that in the African situation, the local languages usually serve as the matrix language while English or French is the embedded language (1989: 157). Similarly, Lahlou (1991) maintains that the matrix language is always the speaker's mother tongue.

1.3.3.4 Joshi's proposal

Joshi's papers, which are partly a continuation of Sridhar & Sridhar's (1980) work, come closest to designing a matrix language model and it is here that the terms 'matrix' and 'embedded language' were introduced.⁸ Joshi stresses the asymmetric roles of both languages but gives no definition of the matrix language other than that "speakers and hearers usually agree on which language the mixed sentence is 'coming from'" (1985: 191). He cites evidence from psycholinguistic research on speech errors, lexical decision and aphasia to support the distinction between what he calls closed class and open class items. He also points out that inserted embedded language elements must be congruent to matrix language elements they substitute: "I assume a 'correspondence' between categories of G_m and G_e , for example, NP_m corresponds to NP_e (..)" (Joshi, 1985: 192, where G stands for 'grammar' and $_m$ and $_e$ for 'matrix' and 'embedded language'). Finally he presents a constraint on closed class items (Joshi, 1985: 194):

Closed class items (e.g. *determiners, quantifiers, prepositions, possessive, Aux, Tense, helping verbs*, etc.) cannot be switched

Unfortunately, this formulation is far from precise. From Joshi's clarifications it becomes clear that what is meant is that closed class items are not inserted on their own, but that they can occur in the company of an embedded open class item. So Joshi allows for the insertion of nouns and noun phrases consisting of a noun and a determiner, but not for the insertion of single determiners (1985: 194).

1.3.3.5 Myers-Scotton's Matrix Language Frame model

Joshi's ideas were worked out by Myers-Scotton and Azuma (1990), and later by Myers-Scotton alone in the MATRIX LANGUAGE FRAME (MLF) model. After having worked extensively on the sociolinguistic aspects of codeswitching with data from fieldwork in Kenya and Zimbabwe, Myers-Scotton started to develop this model in several papers since 1990, culminating in *Duelling Languages* (1993b). Myers-Scotton and Jake are still working on the revision and refinement of the model, but

⁸ The earlier unpublished papers by Joshi and Doron were not available to me. This outline is based on Joshi 1985, which is a slightly revised version of Joshi 1982.

in this introductory discussion I will focus on her 1993 book and the 1995 article by Myers-Scotton & Jake.

Myers-Scotton took over the terms matrix language (ML) and embedded language (EL) from Joshi (1985) but replaced his open and closed class items by different categories called SYSTEM MORPHEMES and CONTENT MORPHEMES. System and content morphemes are defined using certain criteria which will be discussed below. Following Joshi, Myers-Scotton's ideas are heavily inspired by psycholinguistic research findings and models of speech production and her explanation of code-switching phenomena is located within this framework rather than in formal grammar theories. As she formulates it, "the MLF model calls on a suprasyntactic level to motivate its constraints" (1993b: 82).

The model distinguishes three kinds of constituents: 1) ML + EL CONSTITUENTS, 2) ML ISLANDS and 3) EL ISLANDS. ML and EL islands must be constituents well-formed according to ML and EL grammar respectively (1993b: 77-8). Focussing on the mixed constituents, she formulates the ML HYPOTHESIS:

As an early step in constructing ML + EL constituents, the ML provides the morphosyntactic frame of ML + EL constituents (1993b: 82)

From this general hypothesis, two testable hypotheses follow:

The MORPHEME-ORDER PRINCIPLE: In ML + EL constituents consisting of singly occurring EL lexemes and any number of ML morphemes, surface morpheme order (reflecting surface syntactic relations) will be that of the ML.

The SYSTEM MORPHEME PRINCIPLE: In ML + EL constituents, all system morphemes which have grammatical relations external to their head constituent (i.e. which participate in the sentence's thematic role grid) will come from the ML (Myers-Scotton, 1993b: 83)

The System Morpheme Principle blocks EL system morphemes in mixed constituents. The proviso that only system morphemes "which have grammatical relations external to their head constituent" must come from the ML is intended to explain the occurrence of EL derivational morphology and EL plural nouns in codeswitching (cf. Myers-Scotton, 1995: 79; Myers-Scotton & Jake, 1995: 998). The proviso leads to problems in the case of EL plurals, however. Where the matrix language is a Bantu language, as in Myers-Scotton's data, double plural marking from both the EL and the ML is a common phenomenon. Bantu plural markers also indicate noun class. Certain noun class prefixes allow for 'zero allomorphs' so that double morphology can be assumed even if it does not surface (as in the NP *Ø-book-s z-angu* in (31)). In these cases Myers-Scotton argues that only the ML affix is 'relevant', as it is formalized in her Double Morphology Principle (1993b: 133). Crosslinguistically, however, EL plural nouns that function as ML plurals are at least as common as

double plural morphology. These EL plurals, whether they occur as EL islands or within mixed constituents, typically trigger agreement where appropriate according to the ML grammar. Consider (32) where the Dutch suffix *-en* marks plurality in the NP *duk artikelen* “those articles”. The ML determiner *duk* and the Object suffix *-hũm* display number agreement. In the same way, in (33) the French definite article *les* marks plurality in the mixed Subject NP *ga ʃles profs* “all the lecturers” and the ML verb agrees in number with the Subject. (More examples are discussed in Chapter 2, p. ?.)

- (31) leo si-ku-come na ø-book-s z-angu
 today 1S·NEG-PAST·NEG-come with CL10-book-PL CL10-my
 “Today I didn’t come with my books.”
 Swahili/English (Myers-Scotton, 1993b: 80)
- (32) duk *artikel-en*, ila bgi-ti t-teržem-hũm, *is echt moeilijk*
 DEM·PL article-PL if want-2SG 2-translate-3PL is really
 difficult
 “Those articles, if you want to translate them, that’s really difficult.”
 Moroccan Arabic/Dutch (Hocine)
- (33) ga ʃ *les profs* ka-y-xerri-w-ek (..)
 all DEF·PL teacher ASP-3-make·shit-PL-2SG
 “All the lecturers make you shit (..)”
 Moroccan Arabic/French (Wernitz, 1993: 227)

If there is just an EL system morpheme marking plurality in a mixed constituent, it cannot be said to be syntactically irrelevant. This EL system morpheme challenges the System Morpheme Principle and the proviso stated in the Principle cannot solve the problem. As an alternative solution the MLF model may classify EL plural nouns as (internal) EL islands (see below).

Definition of the Matrix Language in the MLF model

Apart from this, the model seems to be supported by various data sets. Of course, the validity of the model eventually depends on the definition of the matrix language and the definition of system morphemes as opposed to content morphemes. In an attempt to avoid circularity, Myers-Scotton initially (1993b) decided to identify the matrix language independently of the structural role it plays. For this reason she proposed to count morphemes in a ‘relevant stretch of discourse’:

The ML is the language of more morphemes in interaction types including intrasentential codeswitching (Myers-Scotton, 1993b: 68)

This definition is problematic since it leads to discussions about what discourse samples would be valid and which would count as morphemes. Also it does not really avoid the supposed circularity because in languages where system morphemes normally outnumber content morphemes, it is precisely the structural role of the ML which leads to the correct outcome. (Conversely, if the (matrix) language is a language in which content morphemes normally outnumber system morphemes, this method would yield the wrong result if all content morphemes in a stretch of discourse were embedded!) One wonders how this morpheme frequency criterion can be reconciled with Myers-Scotton's acknowledgement that, at least theoretically, the ML may change within the same sentence (1993b: 70). Problems arise as soon as both languages seem to function in turn as the matrix.

One of the later versions of the MLF model (Myers-Scotton, 1995: 77) mentions some additional criteria which are no less problematic: the ML is the language that is the unmarked or expected choice, and "the language which the subjects engaged in CS will identify themselves as the main language being used". Both these criteria and the morpheme frequency criterion make the ML and thus the grammatical analysis of a syntactic structure dependent on external variables. Yet it seems straightforward to attribute a grammatical structure to one language rather than the other on the basis of the properties of the structure at hand. This yields an ML that is identified by its role in structuring the clause, as proposed by Klavans (1985), Nishimura (1986), Schmid (1986) and Pandit (1986). The ML in mixed clauses will often prove to be the predominant language in the wider discourse context. However, there is no motivation to identify the language of the morphosyntactic frame on the basis of this context. If one chooses to infer the ML from the structure of the clause, this means that the ML is at first determined post hoc, but it does not preclude the formulation of generalisations concerning, for instance, the interrelations of the ML and sociolinguistic variables at a later stage. Indeed, during the past few years Myers-Scotton has moved increasingly toward this position. On the definition of the ML Myers-Scotton & Jake (1995) write: "The first criterion is this: the ML is the language that projects the morphosyntactic frame for the CP that shows intrasentential code switching" (1995:983), where the crucial word is *first*. Myers-Scotton (1997) gives the following tripartite definition:

- [1] The ML is the language projecting the morphosyntactic frame for the entire CP which shows intrasentential CS. Two other parts of the definition have to do with morpheme frequency: (a) [2] Generally, the ML is the language contributing more morphemes in a sample of discourse-relevant intrasentential CS (minimally two contiguous CPs, either from a single speaker or from an adjacency pair produced by two speakers); (b) [3] It is also generally the language of more morphemes in the discourse as a whole, including monolingual stretches (Myers-Scotton, 1997: 223, numbers added)

This definition is problematic in three ways. Firstly, the question arises as to how the ML is determined in cases where Myers-Scotton's three criteria do not converge. Is the MLF model upheld as long as one of these criteria points to the right ML? In that case the predictions of the model are hard to falsify but the predictive value diminishes accordingly. On the other hand, if there is a hierarchy among these criteria, why not confine the definition of the ML to the decisive criterion and reduce the other two to likely correlates? Secondly, the hedges in criteria 2 and 3 and the open-ended definition of 'a sample of discourse-relevant intrasentential CS' and 'the discourse as a whole' make these criteria hard to work with. Thirdly and most importantly, the most firmly stated criterion 1 seems tautological. It relocates the definition of the ML to the task of recognising 'the language projecting the morphosyntactic frame for the entire CP'. Setting the frame is, after all, what one expects from a matrix language. But how do we identify this language? Myers-Scotton & Jake (1995: 983) provide an answer: "A major aspect of this criterion is operationalized as the *morpheme order* and *system morpheme* principles of the MLF model." This completes the circle. The circularity Myers-Scotton was so anxious to avoid in *Duelling Languages* (1993b) is now reintroduced in full. The ML is inferred from the morphosyntactic make-up of the mixed constituent. More specifically, it is the language to which the system morphemes and the morpheme order can be attributed. I agree that the ML can only be established on the basis of the morphosyntactic make-up of the matrix structure itself, but then it can no longer be presented as a hypothesis that the ML projects the morphosyntactic frame. Rather, the prediction should be that the morphosyntactic frame can be attributed to a single language.

Another aspect of 'the language projecting the morphosyntactic frame for the entire CP' might be the order of constituents within the CP, such that the ML is inferred from the constituent order along the lines of Sridhar & Sridhar (1980) and Nishimura (1986). However, it seems that according to the MLF model, the ML does not necessarily coincide with the language of the constituent order, as the word *may* in the following citation suggests: "Some aspects of EL islands may be determined by the ML, for example, their position in the larger CP" (Jake & Myers-Scotton, 1997: 26). Moreover, Myers-Scotton and her colleagues continue to insist that "one has to examine a discourse-relevant sample (at least two sentences) in order to determine the ML" (Myers-Scotton et al., 1996: 10 n.2).

System and content morphemes in the MLF model

Concerning the definition of system morphemes, again three criteria are identified: [+/- Quantification], [+/- Thematic Role-Assigner] and [+/- Thematic Role-Receiver]. Categories that 'point to particular individuals' receive the feature [+ Quantification] and are system morphemes. For instance, tense and time adverbs "pick out individual events or subsets out of sets of possible events" (1993b: 100 n. 17). Categories with a minus setting for [Quantification] and a plus setting for either [Thematic Role-Assigner] or [Thematic Role-Receiver] are content morphemes. Prototypical thematic role assigners are verbs; nouns are prototypical thematic role receivers (1993b: 98-

101). The system is more complicated than just described, however this will suffice for a brief discussion.

Differentiating system and content morphemes in a principled way that is valid for all languages is an extremely complicated task. I will only mention some major objections to the criteria advanced by Myers-Scotton. The criteria must not be manipulated such as to make certain word classes fit into the right category. Firstly, the argument that pronouns and descriptive adjectives are potential thematic case receivers since they are dominated by the category NP (1993b: 126) raises the question as to whether thematic roles are assigned to constituents or content morphemes. Elsewhere, Myers-Scotton argues that content morphemes are distinguished from system morphemes because the former are potential thematic role assigners or receivers (1993b: 100). If one maintains that thematic roles are assigned to content morphemes rather than to constituents, then it is not clear why descriptive adjectives are thematic role-receivers.⁹ Besides, descriptive adjectives can be said to ‘point to particular individuals’ and pick out an individual item or a subset, and thus bear the feature [+ Quantification] on a par with determiners. Secondly, while free form pronouns are thematic role receivers and thus content morphemes, clitic pronouns are system morphemes because these are actually “agreement particles which are co-indexed with a null NP head” (1993b: 126). Although it helps to uphold the MLF model, this analysis of free and clitic pronouns appears arbitrary. The syntactic representation of pronoun clitics is a matter of much debate and, according to some syntacticians, their function is precisely to receive thematic roles (Borer 1986). Thirdly, in the more recent versions of the model, conjunctions, which do occur often as single embedded forms, are said to be thematic role assigners at the discourse level (Myers-Scotton et al., 1996: 13; Myers-Scotton & Jake, 1995: 984). Under the present criteria of the MLF model, the status of derivational morphemes and various kinds of adverbs other than time adverbs remains obscure.

With respect to the latter, the prediction that they do not occur as singly embedded morphemes, which follows from their classification as system morphemes, is not corroborated by the data. Lehtinen (1966: 227) lists a number of embedded English time adverbs in her American Finnish/English corpus (cf. (34)) and counter-examples occur in other CS data as well. In (35) the application of the Dutch ‘verb-second’ rule, according to which the finite verb remains in the second position in declarative main clauses, identifies the Moroccan Arabic time adverb *ḡedda* as the first constituent in the Dutch clause.

- (34) minä näe minu äiti oikeen *seldom* tiäkkö
 I see my mother really seldom you-know

⁹ One might just as well claim that adjectives assign thematic roles to the nouns they modify (e.g. the role of EXPERIENCER for an adjective like *angry* in *the angry young men*).

“I see my mother really seldom, you know.”
Finnish/English (Lehtinen, 1966: 227)

- (35) *ġedda kan ik niet*
tomorrow can I not
“Tomorrow I can’t [come/go].”
Moroccan Arabic/Dutch (Nortier, 1990: 128)

Another flaw of the MLF model concerns the definition of ML or EL constituents, also called islands. The model does not provide for the possibility of single morpheme constituents since “all islands must be composed of at least two lexemes/morphemes in a hierarchical relationship” (1993b: 138). ML and EL islands “must show internal structural-dependency relations” (1993b: 78).¹⁰ Consequently, the time adverb *ġedda* “tomorrow” in (35) cannot be interpreted as an EL island, even though Dutch syntactic rules identify it as the first constituent of the clause. (Note that the classification of *ġedda* as an EL island would have saved the model from a counter-example.) Likewise, Muysken & De Rooij (1995:1047) note that *somebody* in (36) cannot be an EL island in the MLF model.

- (36) *you didn’t have to worry que somebody te iba a tirar con cerveza o una botella or something like that*
“You didn’t have to worry that somebody was going to throw beer or a bottle at you or something like that.”
Spanish/English (Poplack, 1981: 170)

And indeed, Myers-Scotton (1993b: 127) analyses *somebody* in the above example as an embedded content morpheme in an ML + EL constituent. It is not clear from the text which ML + EL constituent Myers-Scotton envisages here, but it is probably a higher order constituent such as IP or CP.

Single morpheme constituents do not fit in to the MLF model because the model excludes the embedding of single system morphemes. Since EL islands may contain EL system morphemes, the possibility of single morpheme islands would create an escape hatch from the System Morpheme Principle: EL system morphemes that, according to this principle, should not surface in ML + EL constituents could still surface under the label of EL islands. After all, each single morpheme is ultimately a constituent on the lowest level of the structural hierarchy.

¹⁰ If there are no single morpheme EL or ML constituents in the MLF model, single morpheme ML + EL constituents seem even more unlikely. ML + EL constituents “typically consist of morphemes from both the ML and the EL” (1993b: 77).

The exclusion of single morpheme constituents from the MLF model does not take into account distributional facts that are traditional criteria for constituency. In terms of distribution, pronouns like *somebody* in (36) have more in common with NPs than with nouns. But the architecture of the MLF model requires that *somebody*, and other pronouns, are classified as content morphemes, whereas more complex NPs are recognised as EL islands (cf. Myers-Scotton, 1993b: 144).

Congruence in the MLF model

Finally, cross-linguistic congruence of categories was not initially addressed as a prominent factor in Myers-Scotton's model. It was more salient in Joshi (1982, 1985). Congruence did have a place, however, in Myers-Scotton's BLOCKING HYPOTHESIS which further limits the insertion of EL content morphemes in ML + EL constituents:

In ML + EL constituents, a blocking filter blocks any EL content morpheme which is not congruent with the ML with respect to three levels of abstraction regarding subcategorization (Myers-Scotton, 1993b: 120)

These three levels are: 1) content/system morpheme status, i.e., an EL content morpheme can only occur when its ML counterpart is also a content morpheme; 2) thematic role assignment; 3) discourse or pragmatic functions (1993b: 121). This definition of congruence also shows that Myers-Scotton assumes that there is an ML counterpart lemma behind each embedded content morpheme (1993b: 118-9). Myers-Scotton (1993b: 49) follows Levelt in the definition of the lemma as "the non-phonological part of an item's lexical information" (Levelt, 1989: 6). Each lemma contains the pragmatic, semantic, and grammatical information of a lexical item. Although congruence had a rather marginal place in the 1993 version of the MLF model, Myers-Scotton, like Poplack, has begun to concentrate more on this facet of codeswitching during the last few years.

Myers-Scotton & Jake (1995) work out the matching of EL and ML lemmas in detail. Moreover, this publication shows that the conception of congruence and the consequences of insufficient congruence have developed significantly since the 1993 model. In Myers-Scotton (1993b) the emphasis was on congruence as a potential inhibiting factor; now the emphasis is on how congruence permits the embedding of EL content words. The embedding of an EL content word still involves both an EL and an ML lemma. The ML lemma sends directions to what is called the Formulator, after Levelt (1989). The Formulator is the part of the speech production model that calls frame-building morphosyntactic procedures. "How the EL morpheme may appear in a CP framed by the ML depends on the extent to which there is congruence between its lemma and an ML counterpart in the mental lexicon" (Myers-Scotton & Jake, 1995: 1017). Again, congruence is considered on three levels: 1) lexical-conceptual structure, 2) predicate-argument structure, and 3) morphological realization patterns. Insufficient congruence between the selected EL lemma and any

ML lemma leads to a number of COMPROMISE STRATEGIES. The most important compromise strategies are a) the production of ‘bare’ EL forms that lack both the obligatory ML morphosyntactic marking and EL marking, b) the production of so-called EL islands and c) periphrastic *do*-constructions (Jake & Myers-Scotton, 1997).

The quintessence of the MLF model is that each individual EL lemma is matched for congruence against an individual ML counterpart lemma. If the EL lemma turns out to be congruent, it will surface as an EL content morpheme. One might object that many EL content morphemes do not have a matching lemma in the ML, and that the motivation for their insertion is precisely the fact that they fill a lexical gap in the ML. But in the case of lexical gaps, the authors claim, the EL lemma is checked against characteristic lexicalization patterns of the ML.

The most important critique of this approach to congruence is that (lack of) congruence is largely determined post hoc. Moreover, creating a division between EL words that fill lexical gaps and others that do not contributes little to the model, while generating serious practical and theoretical problems. To start with the practical problems, in most cases it is impossible to determine whether an EL lemma has been matched against an ML counterpart and embedded as an EL content morpheme, or is following a characteristic lexicalization pattern of the ML (as do words that fill lexical gaps). The apart procedure for lexical gaps becomes apparent only in comparison with non-congruent EL lemmas, preferably from the same word class (e.g. nouns, or verbs). But nowhere in the data do we observe a morphosyntactic divide between EL morphemes filling lexical gaps in the ML and EL lemmas having incongruent ML counterparts. In Turkish/Dutch, for instance, Dutch verbs are embedded in Turkish periphrastic *do*-constructions with the Turkish verb *yap*- ‘to do’ (Backus, 1992, 1996b).

- (37) ama ben onu hep *uitschrijv-en* yap-acağ-ım
 but 1SG 3SG·ACC all transcribe-INF do-FUT-1SG
 “But I will transcribe it all.” Turkish/Dutch (Backus, 1992: 77)

All Dutch verbs in Turkish are integrated in this way, and consequently, Jake & Myers-Scotton (1997) contend that all Dutch verbs are incongruent with Turkish verbs because of differences in the way ‘the lexical-conceptual structure is mapped onto the predicate-argument structure’ in each language. (That is, Turkish verb complements are overtly case-marked, whereas Dutch verb complements are marked by word order and prepositions.) While this may be true, it is noteworthy that the ‘compromise strategy’ applies to all embedded verbs, regardless of whether or not they fill lexical gaps in the speakers’ Turkish. Words filling lexical gaps are not moulded into ‘characteristic lexicalization patterns of the ML’, after all. We must conclude that (lack of) congruence pertains to lexical categories like Verb rather than individual lemmas. The adaptation of the concept of congruence in the MLF model in this sense would seriously upset the model and the reasonings built on the presumed incongruence of individual EL lemmas.

Finally, there may be other ways to account for the compromise strategies that are claimed to symptomise insufficient congruence at some level. Many bare forms occur systematically across a class of inserted words. These are probably due to unproductivity of the relevant ML morphological procedure, see section 4.2 below. In the MLF model, inserted constituents or EL islands are associated with a) formulaic expressions and b) certain syntactic positions (Myers-Scotton, 1993b: 144-7). In Chapter 11 of this study, EL constituents will be further associated with c) topicality and d) repetition. Since there are many competing explanations for the emergence of EL islands it is difficult to invoke an EL island as evidence for incongruence of a particular EL lemma. It furthermore is, possible to attribute the regularities in the presence or absence of EL constituents to factors that are labelled ‘congruence’ at constituent level along the lines of Bautista’s (1975, 1980) model. In the case of periphrastic *do*-constructions, two factors other than congruence in the sense of the MLF model have to be taken into account. Firstly, periphrastic *do*-constructions are a pervasive feature of various (matrix) languages, to the extent that they will be used to incorporate foreign verbs from any language, irrespective of how the EL ‘maps the lexical-conceptual structure onto the predicate-argument structure’. The periphrastic *do*-construction has been observed for many Turkic and Indo-Iranian matrix languages in various language contact situations, for instance. For other matrix languages the use of a periphrastic *do*-construction is less obvious. Secondly, there are some indications that certain sociolinguistic parameters of language contact encourage the use of a periphrastic *do*-construction in some of the languages where such a construction is less obvious. I will discuss this in Chapter 11, section 1.2.

Further comments

A few more general comments on Myers-Scotton’s model can be added. The model concentrates on content morpheme insertion in mixed ML + EL constituents, giving less attention to EL constituents (EL islands). But especially embedded units that are larger than a single content morpheme but smaller than a constituent receive somewhat stepmotherly treatment. We have already touched upon the treatment of EL plural nouns in the MLF model. Other examples include EL nouns plus an EL determiner or adjective in a larger matrix language NP, as in (33) above and (38) and (39):

- (38) *dak la chemise*
 DEM DEF-F shirt
 “That shirt” Moroccan Arabic/French (Bentahila & Davies, 1983: 317)
- (39) *ka-y-dir-u dik technisch-e school*
 ASP-3-do-IMP-PL DEM technical-AGR school
 “They go to that Polytechnic School ..” Moroccan Arabic/Dutch (Warda)

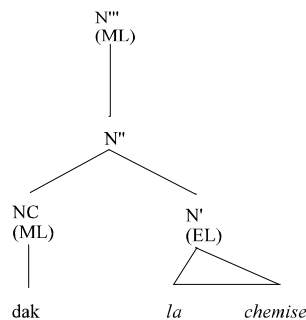
Myers-Scotton calls these cases INTERNAL EL ISLANDS, i.e. EL islands within ML + EL constituents (1993b: 151). Although internal EL islands are frequent in perhaps all data sets, they are treated as extraordinary cases in the MLF model. Myers-Scotton proposes an analysis of EL islands in terms of layered ‘projections’ under X-bar theory.¹¹ I have no quarrel with the notion of internal EL islands, however it undermines the authority of the System Morpheme Principle. Surely, one may decide to classify the EL plural nouns in (32) and (33) as internal EL islands instead of EL content morphemes, but the model makes no predictions concerning internal EL islands, except for the claim that EL islands in general result from the accessing of EL lemma which has no congruent ML counterpart.

Concerning this point Hasselmo’s Ordered Selection model, which identifies several levels between content morpheme and constituent, appears more elegant, covering the various forms of codeswitching (content morpheme insertion, derivation, plural nouns, internal EL islands and EL islands) in a uniform manner. The layered projections in X-bar theory are obviously partly another formalization of what Hasselmo calls levels.

1.3.3.6 Selectional properties of the ML (Schmid 1986)

Finally, I would like to draw attention to Schmid’s 1986 article in which she addresses the selectional properties of the ML. In an ML sentence ML selectional properties apply, that is to say, the ML determines which obligatory constituents make up the sentence structure and which optional constituents may occur. Schmid approaches the issue from the opposite end by identifying the ML by its selectional properties. Or, as she puts it, “if the sentence is grammatical, and contains rules present in only one of the two involved languages, the language containing these rules must be the matrix language” (1986: 74). The example she uses to illustrate this principle concerns the admissibility of resumptive pronouns in certain contexts of extraction in Swedish and English. I will not go into the syntactic intricacies of extraction from embedded clauses, but example (40) will illustrate the main point.

¹¹ Thus example (4) *dak la chemise* is represented as a binary branching configuration with a top node N''' (N triple bar), which is then ‘projected’ as N'' (N double bar)



with two sister nodes: the noun complement or Specifier *dak* and an N' (N-bar) node. This is the internal EL island *la chemise* (Myers-Scotton, 1993b: 153).

- (40) a. The man who I think ____ saw Mary
 b. * The man who I think that he saw Mary
 c. Mannen som jag tror ____ såg Marie
 d. Mannen som jag tror att han såg Marie
 e. Mannen som jag tror *that he saw Mary*
 (Swedish/English, constructed examples, Schmid, 1986: 76)

First note that (40)b is not grammatical while its word for word equivalent in Swedish, (40)d, is. Now according to Schmid, (40)e is judged to be grammatical because the matrix language is Swedish and the embedded complement clause *that he saw Mary* is equivalent to the Swedish complement *att han såg Marie*. The fact that a parallel construction is not possible in English shows that Swedish must be the matrix language in (40)e.

Schmid's example is based on constructed sentences and grammaticality judgments by a panel of bilingual speakers. Obviously, it would need confirmation from naturally occurring data. The point Schmid makes perhaps more clearly and more explicitly than any other author, is that the matrix language manifests itself not only in providing constituent order and inflection, but also in the specifications that hold for each slot in the syntactic frame.

1.3.3.7. Some general remarks on matrix language models

The above has shown that there are significant advantages in the matrix language, or insertional, approach. However, this approach is at times problematic and a certain amount of scepticism is justified. The main objections to ML models in general may be summarised as follows:

Firstly, the matrix language approach is only concerned with syntax and morphology. It obviously has little to contribute to the discussion of phono-logical or discourse related aspects of codeswitching.

Secondly, there is the problem of identifying the matrix language in an unambiguous and non-circular way. No matter what criteria are applied, intricate exceptions and counterexamples will always remain. It is possible that in some cases different matrix languages have to be assumed for different levels of speech production, thus rendering the approach quite complicated. Besides, as Muysken (1995: 180) notes, the model becomes less attractive in cases where a number of adjacent EL constituents have to be assumed. Since the matrix language is strongly associated with function morphemes, the differentiation of function and content morphemes is another predicament for the matrix language approach.

Finally, the insertion of all kinds of constituents and internal constituents (Myers-Scotton's internal EL islands) is subject to regularities and restrictions that call for ampler investigation. At present these regularities are mostly attributed to the congruence factor, but eventually more sophisticated explanations will be required for at least some of these phenomena.

1.3.4 Conclusion

Comparing the government model with Myers-Scotton's matrix language model, Muysken in his 1995 paper summarizes that the former specifically excludes functional elements from being relevant governors, whereas the latter focuses on functional elements as governors. He then concludes: "It is fair to say that this latter option must be closer to the truth" (Muysken, 1995: 188). Indeed, as I have pointed out in 1.3.2, the government models that are most successful in capturing the regularities in codeswitching are those that allow for lexical and constituent insertion, and identify function words as governors. These are the government models that most resemble the Matrix Language Frame model in its prediction that in mixed constituents, function morphemes and word order are basically from one language - the matrix language. I have also demonstrated that the weakness of these models lies exactly where they fail to separate function and content morphemes, namely when the inserted content morpheme is a governor itself. Transitive verbs are governors in any current definition of government and consequently the insertion of a transitive verb in an otherwise ML structure is mistakenly ruled out by the models of Pandit (1990), Santorini & Mahootian (1995), Halmari (1993) and Muysken (1995).¹² Thus it seems safe to conclude that, as a single principle regulating codeswitching behaviour, the content/function morpheme distinction constitutes a generalisation that covers more data than the government principle. However, whereas the advocates of the government model struggle with the identification of the relevant governors, the adherents of the matrix language model face the intricacies of defining the categories of content and function morpheme. This matter will be pursued in the next chapter.

The advantage of the government model over Myers-Scotton's MLF model is that it allows the description of insertions without the necessity of identifying a matrix language. Also, taking as its point of departure the idea that each governing element creates its own matrix structure, the government model can easily deal with successive layers of insertion. Consider the Object NP *la personnalité dyal quelqu'un* in the French clause in (41). [*dyal* [*quelqu'un*]] can be analysed as an Arabic PP that has been inserted under structural equivalence; the preposition *dyal* projects its own constituent structure, in which French *quelqu'un* is inserted.¹³ The governing nature of the inserted preposition can be illustrated more convincingly in the second example

¹² Note that to classify EL verbs such as *pretend* in (20) as borrowings is not a desirable option. It would lead to all singly occurring, governing EL verbs being categorically classified as borrowing, with the result that the borrowing/codeswitching distinction would become meaningless for this word category (cf. section 1.4).

¹³ Alternatively, one may regard *ça reflète la personnalité dyal quelqu'un* as an inserted IP within an Arabic CP governed by the foregrounding element *ra-h*. Such an analysis is proposed by Myers-Scotton and her associates for Arabic discourse emphatic pronouns, see the discussion in Chapter 2.

(42), where *mit piriñ* clearly displays the word order of the German preposition since Turkish only has postpositions and postnominal case-markers (e.g. Standard Turkish *piriñ-le* “with rice”).

- (41) *ra-h ça reflète la personnalité dyal quelqu'un*
 PRES-3M this reflects the personality of someone
 “This reflects someone’s personality.”
 Moroccan Arabic/French (Wernitz, 1993: 340)

- (42) *pilav mit piriñ yap-1l-1r*
 boiled·rice with unboiled·rice make-PASS-AORIST
 “Boiled rice is made with unboiled rice.”
 German/Turkish (Treffers-Daller, 1995: 254)

Note that examples such as (41) and (42) are difficult to deal with in the classical Matrix Language Frame model, due to the difficulties in defining the matrix language. The advantage present in the government model can be found in any type of grammatical analysis that recognises an ordering of constituents in a hierarchy of structural layers without the necessity of identifying a governing morpheme that creates the constituent structure.¹⁴ (See for instance the informal ‘levels’ in Hasselmo’s Ordered Selection). In this vein the matrix language would need to be defined more broadly as “the language which projects the grammatical frame for the unit showing intrasentential CS” (Myers-Scotton et al., 1996: 16), where ‘unit’ is the crucial term, as opposed to “the language that projects the morphosyntactic frame for the CP” (Myers-Scotton & Jake, 1995: 983; cf. Myers-Scotton, 1997: 223). That is, we need to recognise that various lower and higher-order morphosyntactic structures like the CP and its constituents can function as a matrix in which elements from the other language are embedded. Thus the ML is not predicted by an independent criterion, but inferred from the make-up of the matrix structure itself. The only claim to be made by the model is, therefore, that the morphosyntactic features of the matrix structure can be attributed to the rules of one language, rather than to the overlapping parts of the grammars of both languages, or to a third grammar specific to the CS variety. In many languages, grammatical structure is marked by word order and/or function morphemes, so that these aspects will be central in identifying the ML. In the next chapter, this method will be presented in more detail.

While content word insertion and constituent insertion seem to be major principles regulating codeswitching it is by no means the case that any kind of content word

¹⁴ In Levelt’s (1989) psycholinguistic speech production model, it is argued that constituents are assembled by independent phrase structure components: “the lemma’s syntactic category calls a categorial procedure, namely a building instruction for the phrasal category in which the lemma can fulfill the function of head. This holds for the lemmas of types N, V, A, P” (1989: 238).

or constituent can be freely inserted. For this reason any model of codeswitching requires the concept of congruence (structural equivalence) in order to account for the regularities in insertional patterns that are not covered by the function/content morpheme distinction. In the first place, the notion of congruence explains that inserted elements occupy particular slots in the matrix language frame, for instance, that the German PP *mit piring* in (42) occupies the slot for the instrumental complement in the Turkish verb phrase. In the second place, the concept of congruence is invoked to explain why certain kinds of insertions are frequent, while other kinds of insertion do not occur at all. Congruence is at least an implicit factor in every approach that recognises insertion. It is made explicit in the papers of Sridhar & Sridhar (1980), Joshi (1982, 1985), and Schmid (1986). The concept takes a more concrete shape in Bautista's Table of Phrase Structure Equivalences and Myers-Scotton's Blocking Hypothesis. The former deals with congruence on the phrase level and the latter with the morpheme level, so that the contributions of Bautista and Myers-Scotton are complementary.

Bautista applies the congruence factor to CATEGORIES of embedded phrases whereas Myers-Scotton believes that every embedded content morpheme is individually checked for its congruence with an ML counterpart - except in the case of EL words that fill lexical gaps in the ML. Relating congruence to lexical and phrasal categories seems preferable as it generates broader generalisations and reduces the risk of invoking the congruence factor as an ad hoc explanation (cf. Schmid, 1986: 76 for a similar argument).

Exactly what constitutes congruence is not entirely established. So far it can only be observed post hoc: bilingual speakers turn out to insert certain categories of embedded language words and constituents in the morphosyntactic positions of corresponding categories from the matrix language. The first step in uncovering the meaning of congruence consists in the design of a matrix language model. Subsequently, the study of insertion patterns in various sets of CS data will allow us to make generalisations as to what factors appear to favour congruence. For this purpose it will be useful to distinguish between embedded content words, inflected content words and constituents. Since individual codeswitchers happen to differ with respect to the word and constituent types they insert, congruence appears to be at least partly idiosyncratic. A number of factors are probably involved in the individual speaker's perception of a congruence relation between categories from one language and the other: universality, as opposed to language-specificity of lexical categories; formal or functional similarity of ML and EL categories; and factual or perceived overall relatedness of the languages involved. In addition, the types of insertion that occur depend on the sociolinguistic status of the ML in the bilingual speech community (Chapter 11).

As superficial similarity in general contributes to the establishment of congruence, this also includes word order similarity. I concur with Muysken's (1995) interpretation of word order equivalence as a subcategory of structural equivalence (which I take to be synonymous with congruence). I have shown in section 3.1 above that

word order equivalence is not a major principle ordering codeswitching. On the other hand, it is possible that word order equivalence facilitates the establishment of congruence in certain cases. At this point, my impression is that word order differences scarcely affect the insertion of content words, while they do play a role when it comes to embedding certain kinds of constituents. Modifier constituents such as relative clauses, attributive PPs and certain adverbial phrases typically require to be either followed or preceded the head they modify, whereas the placement of other constituents, notably the major arguments of the verb, is subject to pragmatic and syntactic rules from the matrix language. As a consequence, assume that EL modifier constituents would easily occur in situations where the ML and the EL share the same head-modifier order. This remains to be established, however.

In conclusion, congruence is a cover term for a number of very diverse mechanisms, some of which will eventually receive a more accurate interpretation, and some of which may at a later date turn out to be better explained by a completely different principle. I will not pursue the topic of congruence in the present study, but see Boumans (1995b and forthcoming) for a more detailed discussion of congruence between inflected word forms of the ML and the EL. My own proposal for a matrix language model will be outlined in the next chapter, but first I will address the distinction between borrowing and codeswitching.

1.4 The codeswitching versus borrowing discussion

There is a long-standing debate on the differentiation of codeswitching and borrowing (B). I will not attempt to review the huge literature on the distinction between CS and B, but will point out the major issues, with special attention for the literature on Moroccan Arabic. For this discussion we need a neutral cover term in order to speak about ‘words from another language’ without a priori deciding on their status as an instance of borrowing or codeswitching. I will use the term ‘foreign lexeme’ for this purpose. I consider the terms ‘loanword’ and ‘borrowing’ to be equivalents.

Loosely speaking, most authors view B forms as part of the (matrix) language in which they occur, while CS forms are generally considered to be part of the ‘other language’. Although one may feel intuitively as to which language a word really belongs to, many scholars have been, and still are, trying to formalise this difference. The distinction of CS and B has raised much controversy because the category of borrowing is sometimes invoked to explain apparent counter-examples to formulated constraints on codeswitching. By calling them (nonce-) borrowings, singly inserted words, in particular, can be excepted from the applicability of codeswitching rules. Or, as Schatz (1989: 129) puts it, borrowing and nonce-borrowing “often seem to play the role of garbage can designed to throw in data that does not fit neatly defined constraints on codeswitching” (cf. also Giesbers, 1989: 21 for a similar remark). Poplack and her associates, for instance, introduced the notion of nonce-borrowing

arguing that nonce borrowings do not have to comply with the constraints on CS which they formulated, notably the Equivalence and the Free Morpheme Constraints (cf. section 3.1 above). In Myers-Scotton's MLF model, the distinction between CS and B plays a role primarily with respect to function morphemes. According to her model, EL function morphemes cannot be inserted in ML + EL constituents. The French complementizer *que* that occurs in Lingala/French CS (Kamwangamalu, 1989; Bokamba, 1989) would be a counter-example to the MLF model, but here Myers-Scotton suggests that *que* may turn out to be really a B form (1993b: 132, 205).¹⁵

The discussion mainly revolves around the insertion of single words, since most scholars agree on classifying longer stretches of EL material as CS. While there may indeed be different constraints on the insertion of single words as against the insertion of larger stretches, we need to make a principled decision on whether we rate singly inserted words under codeswitching. I have made this decision by sticking to the term codeswitching as defined by Haugen, who unequivocally includes "single, unassimilated words".

Hereafter I will review some of the most commonly applied criteria for the distinction of CS and B, namely the occurrence of a foreign lexeme in the speech of monolingual speakers (section 1.4.1), morphological integration (1.4.2) and frequency (1.4.4). In section 1.4.3 I will pay special attention to the relationship between (phonological and morphological) integration and development over time because the two current views on this relationship are almost opposite to each other. Concluding remarks are presented in section 1.4.5. In addition to the criteria that will be discussed in this section, a number of other criteria are commonly used in the distinction of CS and B, notably phonological integration, the question as to whether the ML has a translation equivalent of the foreign lexeme, and speakers' judgements as to what language a given word belongs to (cf. Schatz, 1988: 132). I will not discuss these criteria here. Although they have intricacies of their own, in many respects the argumentations in the following sections can be extended to encompass these criteria as well. I refer to Heath (1989: 23-5) for an evaluation of the phonological adaptation criterion with regard to MA/French language contact, and to Backus' (1996b) Specificity Continuum for a critique of the idea of translation equivalency (cf. also Chapter 11, p. ?).

It should further be noted that although the discussion concerns the classification of individual tokens of foreign lexemes as either CS or B, many of the criteria that distinguish these categories refer to the status of the type in the bilingual community. This concerns the occurrence of the type in the monolingual speech community, the frequency of the type in a given text corpus, the availability of a 'native' translation equivalent and bilingual speakers' judgment. Only morphological and phonological integration are criteria pertaining to the token at hand.

¹⁵ See section 2.3.1 of Chapter 3 on *que* in Shaba Swahili, another Bantu language of Congo-Kinshasa, former Zaire.

1.4.1 Borrowing, codeswitching, and bilingualism

The definition of codeswitching as the alternate use of two languages implies a certain amount of bilingualism. This distinguishes codeswitching from borrowing to some extent since the latter does not necessarily require a bilingual situation. People may, for instance, know the names of various exotic dishes that are served in restaurants without having any knowledge of the languages these words come from.¹⁶ The use of a foreign word in the monolingual speech community is, in fact, often mentioned among the criteria that set off borrowing from codeswitching. It is probably the most reliable criterion crosslinguistically.

The criterion that loanwords are also used by monolingual speakers relocates the problem to the definition of bilingualism. This definition is no less intricate. And, as Myers-Scotton points out, even a very low degree of bilingualism may suffice to engage in CS. Therefore it is hardly a feasible procedure to label as CS forms only those forms which are never uttered by a person who is, by some definition, a monolingual (Myers-Scotton, 1993b: 193).

But setting aside the difficulty of defining bilingualism, will it help to define CS as opposed to borrowing? We can exclude the possibility that foreign lexemes uttered by monolinguals are CS forms, however there is no implication that all forms uttered by bilinguals, or even the forms produced exclusively by bilinguals, are automatically CS forms. We cannot exclude the possibility that the distribution of a B form is restricted to a community of speakers who all happen to be bilingual. (Many small languages do not even have monolingual speakers, for that matter.)

1.4.2 Morphological integration

In her review of the B/CS debate Myers-Scotton raises a number of objections to the criterion of morphological integration, which she summarises as follows (1993b: 191):

The problem with morphological/syntactic integration as a criterion for B forms versus CS forms is that several different patterns of integration occur, not just one. This survey has pointed out four patterns: (a) not all B forms show complete morphological integration; (b) most CS forms in ML + EL constituents regularly show near-complete morphological integration; (c) when there is incomplete morphological integration, it may characterize both B and CS forms

¹⁶ No one will argue that CS is at stake in such cases. Yet the names of foreign dishes may be identified as CS forms by at least one criterion sometimes invoked in the B versus CS discussion, namely speakers' intuitions as to what language a given word belongs to.

in contrast to indigenous forms; and (d) both forms show syntactic integration.

In addition to the complications Myers-Scotton mentions, I would like to mention other factors that complicate the use of morphological integration as a criterion by which CS and B are distinguished. Firstly, not all morphological processes in a language are equally productive and secondly, productive morphology is more characteristic of some languages than of others. The latter remark becomes relevant when we approach integration from a cross-linguistic perspective. Nortier & Schatz (1988), for instance, compare phonological and morphological integration of single foreign (B or CS) forms in data from five language pairs. Among other things, they find that Dutch foreign lexemes in Moroccan Arabic show the least integration whereas the highest degree of integration occurs with Spanish foreign lexemes in Ecuadorean Quechua. Nortier & Schatz relate the different degrees of integration to the duration of language contact: Quechua/Spanish contact dates back to the 16th century, while the first Moroccans arrived in the Netherlands only in the 1960s. Below I will comment further on the relationship between time depth and phonological and morphological integration. But first let me illustrate just how influential the characteristics of the matrix language are.

This is easily done by comparing Moroccan Arabic and Turkish as immigrant languages in the Netherlands. In the case of MA, hardly any morphological process is applied to embedded Dutch words or constituents, as the corpus description in this study demonstrates. Turkish, on the other hand, is an agglutinative language with a wide range of nominal affixes marking plural, case, possessive and derivation that are perfectly productive with embedded Dutch nouns. Turkish case-marking suffixes, for instance, are at least as productive as the MA prepositions that serve the same functions. Example (43) shows two Turkish suffixes attached to Dutch nouns: the locative case marker *-da* and the agentive suffix *-çi*. The Turkish suffixes, moreover, are subject to vowel harmony, a fact which may also be used to argue that the foreign lexemes are phonologically integrated.

- (43) evet *terras*-da otur-uyor-lar. orada bir *friet*-çi var
 yes pavement-LOC sit-ASP-PL there INDEF chips-AGENTIVE EXIST
 “Yes they sit on the pavement. There’s a chips stand over there.”
 Turkish/Dutch (Backus & Boumans, 1996: 149)

- (44) mnin te-qbeṭ nta __-*pensioen* dyaal-ek (..)
 when 2-get 2M [DEF-]pension of-2SG
 “When you get your pension, (..)” Moroccan Arabic/Dutch (Warda)

In MA, many grammatical categories that are marked by affixation in Turkish are marked differently, by means of an analytic construction (possessive) or by word order and prepositions (case). But the definite article *l-* would be an obligatory affix

in (44) and even this affix is missing. The omission of the definite prefix before Dutch nouns is more or less the rule in MA/Dutch CS (cf. Chapter 5).

It is therefore much easier to argue for phonological and morphological integration of EL words in Turkish than in MA. Moroccan and Turkish migration to the Netherlands both started in the 1960s and the social circumstances of the bilingual groups are rather similar, thus the duration or intensity of language contact does not explain the different integration patterns. Turkish just happens to have very productive morphological processes, while MA does not. In this context let me note that many MA nouns of Berber origin still defy affixation with the definite prefix, even though these loans probably predate the Quechua/Spanish contact in Ecuador (Colin, 1945: 232; Harrell, 1962: 190).

If we take phonological and morphological integration as criteria for distinguishing CS from B, we might conclude that MA has a predilection for CS forms, while Turkish only borrows Dutch nouns, which of course cannot be the purpose of the B versus CS distinction.

1.4.3 The time depth: from borrowing to codeswitching

Phonological and morphological integration of foreign lexemes is associated with borrowing rather than with codeswitching and the highest degree of integration is often found with the oldest foreign lexemes. The reason for this is that the oldest foreign lexemes in use in a bilingual community stem from the earliest stages of bilingualism and in the earliest stages of language contact the speaker of the matrix language has little knowledge of the ‘donor’ or source language (Haugen, 1950: 216-7). So when the speaker introduces words from that language into her own language, she is not able to give a full replication of the donor language’s phonology and morphology. Instead, she will apply her own phonological and morphological systems to interpret and reproduce the foreign language form. Further integration may follow when the foreign lexeme is taken over by other speakers who know even less of the donor language. But in most of the cases that have been studied, exposure to the (economically and/or culturally superimposed) donor language has become more intense over time. In the course of time, the bilingual community makes more use of the donor language, knows that language better, and therefore, more recent foreign lexemes tend to be less integrated than the older ones. Note that this explanation of the correlation between age and integration of foreign lexemes only applies to cases where contact with the foreign language increases over time. Moreover, it applies only within one particular ongoing situation of language contact.

In short, the degree of phonological and morphological integration is inversely proportional to the intensity of language contact and in many cases language contact becomes more intense over time. This mechanism, as described by Haugen for American Norwegian, is primarily responsible for the relationship between time depth and integration (cf. for instance, the discussion of earlier and later Russian loanwords in Asiatic Eskimo in Thomason & Kaufman, 1988: 32-3; Khubchandani, 1968: 183

on Sindhi/English and Van Ness, 1994: 293 on Pennsylvania German). Various scholars hold an almost opposite view on the relationship between time and integration: they see the latter as the outcome of an autonomous process that takes place over time in the borrowing language. This view is represented by Nortier & Schatz in their 1988 article discussed above, and Beaumont (1987), among others. In the same vein, Heath (1989) claims that codeswitching functions as a pathway to borrowing.

There are certainly good reasons to assume such a development towards integration. Loanwords can be subject to language-internal change. In particular, they may become more like native words with similar meanings on the basis of analogy. A process of ‘levelling’ eliminates pointless variation and thereby maximizes the transparency of the language system (Aitchison, 1991: 146). As an example, one could think of nouns that are borrowed in both their singular and plural form such as English *syllabus*, *syllabi*, but receive a ‘native’ plural at a later stage (*syllabuses*) just like the number of Old English pluralisation patterns has been reduced over time. Perhaps Beaumont (1987) provides an example of this. He investigated the plural formation of French-origin nouns in MA. A subclass of loanwords occurs alternatively with two types of MA plural marking. For instance the word *šifur* ‘chauffeur’ may have either the internal ablaut plural *šyafer* or the suffixed plural form *šifur-at*. Beaumont found that ‘unskilled’ women used slightly more suffixed than internal plurals, whereas ‘unskilled’ men had a clear preference for the internal plural. If we agree that internal plurals show a higher degree of integration than suffixed plurals,¹⁷ and if it can indeed be shown that the unskilled men’s linguistic behaviour is innovative while the women’s language is conservative, as is the purport of Beaumont’s article,¹⁸ then this constitutes an example of morphological integration of loanwords as the product of language-internal change.

However, there is a trend in the opposite direction leading to minimal phonological and morphological integration of loanwords and this trend can be at least as

¹⁷ It can be argued that internal plurals are more integrated because they require that the foreign lexeme be analysed into three or four root consonants. Yet I cannot agree with Beaumont’s (1987: 55, 69) assumption that suffixed plurals are formed in analogy with French pluralisation and are less integrated for this reason, nor with the idea that the broken plural is more integrated because it is more ‘genuinely Semitic’. One has to keep in mind that a) suffixation is a common plural type for certain noun classes in MA and b) not all borrowed words are equally liable to be analysed in three or four root consonants, as is required for internal pluralisation.

¹⁸ As Beaumont (1987: 85) states it “le groupe des hommes non-éduqués, de par leur puissance sociologique et leur position pivot (à fonction receveur-donneur), est plus apte à représenter l’avenir de la langue. (..) Le comportement linguistique générale de nos sujets (..) reflète (..) [la] marginalisation des femmes non-éduquées quant à leur comportement linguistique.”

dominant. Integration over time seems plausible when the influence of the donor language diminishes or even stops, as occurs for instance in the case of B forms that are remnants of a substrate language. But when the impact of the (culturally dominant) donor language and culture increases over time and the bilingual population gains more access to that language, as described at the beginning of this section, increasing knowledge of the superimposed language also leads to replacement of older, integrated loanwords with more exact copies of the source word, a process known as DENATIVIZATION (Haugen, 1953: 393-4; Hasselmo, 1969: 70). A successive stage is that the bilinguals start using the superimposed language for communication among themselves: codeswitching. This situation can be observed among linguistic minorities, either indigenous or immigrant, as is the case with former colonial languages in Africa and Asia and presently with English as the language of science and education world-wide. The force of denativization (possibly preceding codeswitching and language shift) overrules the effect of language-internal change on the integration of borrowings. This development, it seems, can be reversed only as a result of a change in power relations.¹⁹ Returning now to the position of French in Morocco, even Heath, who advances the idea that codeswitching is a pathway to borrowing (Heath, 1989), notes that “French-Moroccan Arabic bilinguals now prefer to reborrow the same French terms in shapes closer to the French prototype” (Heath, 1986: 114).

1.4.4 Frequency of foreign lexemes

Several researchers have distinguished CS and B on the basis of the frequency of occurrence of individual word types in a text corpus, e.g. Myers-Scotton (1993b) and Poplack, Sankoff & Miller (1988). For the latter, the degree of frequency mainly distinguishes between what they call ‘nonce borrowings’, which occur only once, and ‘established borrowings’. For Myers-Scotton (1993b) the distinction of B and CS forms is not pertinent to patterns of phonological and morphological integration. For her, the CS/B distinction is important because it determines whether an EL system morpheme may occur in a mixed constituent. Recall that, according to Myers-Scotton, EL lexemes are embedded only if there is a congruent ML lemma (representing the non-phonological part of an item’s lexical information, after Levelt, 1989). This requirement does not hold when the EL lexeme is borrowed, since in that case its lemma is part of the ML lexicon:

¹⁹ To take Morocco as an example once again: when the Spanish territory in the north and the former French ‘protectorate’ were integrated to form the independent kingdom of Morocco, Spanish lost influence to ‘neo-colonial’ French, which, in its turn, is now beginning to lose influence to (American) English as a prestige language.

First, a B form probably has entries in the mental lexicon of both the ML and its parent, the EL. [...] Second, an EL-origin word which is a CS form is accessed in ML +EL constituents through an EL lemma, true, but only if it is congruent with an ML counterpart, while a B form is accessed directly (through its own lemma) (Myers-Scotton, 1993b: 192)

The common assumption among scholars using the degree of recurrence in a text corpus as a criterion is that B forms occur more frequently than CS forms. Of course the frequency of a lexical item depends very much on the concept it encodes, that is, on whether it encodes a concept that is likely to occur in many contexts. Hence, absolute frequency is not a very reliable criterion. A more precise criterion is *RELATIVE FREQUENCY*, that is, the frequency of a word's recurrence relative to that of its indigenous counterpart. The problem with word frequency in smaller data corpora is that it is highly dependent on coincidental circumstances of the recorded discourse, such as the topics under discussion, the speech style and the interlocutors present. Moreover, the repetition of lexical items as a means of creating textual cohesion (Halliday & Hasan, 1976; De Rooij, 1996) increases the frequency of an item once it has been used, irrespective of its being a CS or a B form. This problem can be overcome to some extent by investigating very large corpora, as was done by Poplack, Sankoff & Miller (1988), however, this does not really resolve the question. It is still possible that using the criterion of absolute or relative frequency, a particular lexical item is judged as a B form within a bilingual community under investigation, while at the same time it is a CS form for a subset of speakers in this community, on the basis of the same frequency criterion.²⁰ The problem imposes itself fully when one aims to speculate on the organisation of the mental lexicon of individual speakers on the basis of the frequency of particular lexical items in a large text corpus that includes data from many different speakers (cf. Myers-Scotton, 1993b: 195-202, on English numbers and discourse markers in a Shona/English text corpus).

The next step towards a solution would be to gather a very large text corpus from one individual speaker in order to investigate the B or CS status of particular lexical items in her or his bilingual speech variety. It must be noted, however, that even individual speakers will vary over time with respect to the frequency with which they use particular lexical items.²¹

²⁰ Myers-Scotton acknowledges this problem in note 25 on p. 195.

²¹ I myself, for instance, have been using the English words *codeswitching* and *matrix language* far more frequently in Dutch ML utterances during the years that I was working on this dissertation than ever before and, perhaps, ever after. If their Dutch equivalents return the same very low frequency of occurrence over time, we could say that according to the criterion of relative frequency, *codeswitching* and *matrix language* were B forms in my personal mental lexicon from about 1993 till

In any case, as Myers-Scotton acknowledges herself, relative frequency “will not always prove a workable criterion” (1993b: 204), especially when EL lexemes representing less commonly expressed concepts are being considered.

1.4.5 Concluding remarks

The difference between B and CS is not a fact of nature; it is about how scholars decide to classify a set of data. Indeed, in view of the lack of consensus that generally governs the domain of linguistic terminology, the task of defining the terms one uses becomes imperative. A criterion that is relevant to one research context may not be relevant to another. Much effort is spent on distinguishing CS and B forms in codeswitching research, while the solution seems simple. If one considers CS versus B to be a useful opposition, the problem solves itself in either of the following ways: 1) In some cases the criteria applied identify observably distinct patterns for CS and B in a certain domain, e.g. phonological adaptation. In this case, the distinct patterns will create the distinction between the two, so that there is no difficulty in distinguishing B and CS. 2) The other possibility is the distinction between CS and B that is not found to correspond to observably distinct patterns in a given domain. Thus, one decides that the binary opposition along the lines of the chosen criteria does not do justice to the data under investigation. Accordingly, one may try other criteria or give up the dichotomy altogether. The main problem is that very often the various criteria do not converge to set off B from CS. This means that a distinction on one level does not yield any prediction for another level of analysis. Whether the ‘receiving language’ has a translation equivalent for a given foreign lexeme does not predict much concerning the morpho-phonological integration of the lexeme, nor does morpho-phonological integration predict syntactic integration. Only in cases where different kinds of criteria seem to converge do borrowing and codeswitching become useful labels to distinguish two sets of data.

The effort spent on distinguishing CS and B and the continuing surrounding controversy it is not always proportional to the insights generated by such a distinction. In fact, the labels B and CS run the risk of alienating facts from their explanation. Let me give three examples to illustrate this objection:

1) CS and B are associated with different degrees of morphological integration. However, in many languages morphological characteristics of words are related to word class, or word category, in the first place. Depending on typological characteristics of the ML, the criterion of morphological integration may lead to the classification of all EL verbs as borrowings and all nouns as codeswitches, for example. Now, it will in itself be worthwhile to study the relationship between morphological integration and word category (cf. Pfaff, 1979: 298; Myers-Scotton, 1993b: 197), but the B/CS dichotomy obscures this relationship. Instead of observing

1998; before and after this period, these lexical items would be CS forms.

that a particular EL lexeme is morphologically integrated because it is a verb, one is led to believe it is integrated because it is a B form. Consider Van Ness (1994: 293) who, linking integration to time depth along the lines of Haugen (1950), suggests that English-origin verbs in Pennsylvanian German must on average be older loans because they are more integrated than English-origin nouns.

2) In cross-linguistic comparisons of foreign lexemes from different language pairs variation in phonological and morphological integration is likely to be primarily related to the monolingual ‘receiving’ languages (section 4.3 above). Applying the integration criterion to comparative CS studies may lead to the conclusion that borrowing is characteristic of some languages, and CS of others. Meechan & Poplack (1995) investigate the use of French in two languages of the Niger-Congo family: Wolof as spoken in Senegal and Fongbe spoken in Benin. In both languages adjectival modification and predication is realised by adjectival verbs. Only Fongbe has in addition a small class of ‘true’ adjectives sparingly used in attributive contexts. Meechan & Poplack find that in Wolof/French, French adjectives function like Wolof adjectival verbs, co-occurring with Wolof verbal markers, whereas in Fongbe/French, French adjectives occur mainly as predicates after the Fongbe copula *dò*. (In monolingual Fongbe this copula is almost exclusively used with locative predicates.) The morphological integration and syntactic distribution of French adjectives in Wolof is consistent with that of Wolof adjectival verbs, and much less so with the properties of adjectives in French. The copula construction in Fongbe, on the other hand, resembles the copula construction in monolingual French. Thus, Meechan & Poplack argue, Wolof primarily borrows French adjectives, whereas French adjectives in Fongbe are codeswitches under the Equivalence Constraint. Meechan & Poplack do indicate that the distinct ‘patterns of use’ for French adjectives are due to language-specific properties of Fongbe and Wolof (1995: 189), but here again, I feel that drawing the cross-linguistic differences into the CS/B discussion does not contribute to a better understanding of the language facts.

3) Myers-Scotton (1993b) uses the concept of borrowing in order to link the relative frequency of a foreign lexeme to its status in the speaker’s mental lexicon. In itself it is entirely plausible that there are causal relationships operating in both directions between word frequency and the organisation of the mental lexicon. How these two are related deserves investigation, but this is better achieved without the intermediary concept of borrowing. The label B has no explanatory value and also creates some confusion. After all, B is still commonly identified with phonological, morphological and syntactic integration, while Myers-Scotton rejects these criteria.

Although I have raised objections against the present CS versus B debate among scholars of bilingualism, this does not mean that I am against any form of classification. Whenever the data display different regularities for subsets of foreign lexemes these differences deserve to be noticed and explained. In such explanations, all the criteria invoked in the CS/B debate (integration, frequency, distribution among speakers, speakers’ judgements, time depth, among others, cf. Schatz, 1988: 132 for an overview) can contribute to our understanding of bilingualism.

In my description of the Nijmegen data corpus I will discuss all foreign lexemes without taking into account their classification as B or CS forms. I will indeed point out systematic differences between sets of foreign lexemes, these will be related to word classes and to individual variation among respondents. Differences in bilingual behaviour among respondents will be further related to their sociolinguistic characteristics. I concur with Myers-Scotton (1993b; 1997) that in an insertional approach to CS, all singly occurring foreign lexemes should in principle be syntactically and morphologically integrated to the same extent as the ML categories with which they are congruent. Deviations from this rule do occur, and need to be accounted for. I will further demonstrate that foreign lexemes vary with respect to the amount of EL grammatical information they display in an ML structure. I refrain from attributing this variation to the foreign lexeme's status as B or CS.

Chapter 2

The Matrix Language

The concept of Matrix Language introduced in section 3.4 of Chapter 1 will be developed in this chapter. For ease of reference, and to distinguish it from related approaches, the model advanced here will be called the Monolingual Structure Approach (MSA): each matrix structure is assumed to originate in the grammar of only one language, the Matrix Language. The first section offers a step-by-step presentation of the concept and the considerations that led to the definition of the ML adopted in the present study. The ML defined here is relevant to the morphology and basic syntax of intra-sentential CS to the degree that these are language-specific, but the ML does not make any predictions on syntactic issues that have a strong relation to discourse organisation. (Those aspects of syntax will be examined in Chapter 3.) Section 2 of this chapter deals with the classification of inserted elements according to the MSA, section 3 summarises the predictions of this model and section 4 discusses the main types of counter-evidence to the MSA. A summary is given in the last section.

2.1 Defining the Matrix Language

The concept of a Matrix Language stems from the assertion that a grammatical structure containing elements from two languages can be attributed to the grammar of one of these languages (the ML), rather than to the grammar of both languages, to the overlap of both grammars or to a third ‘codeswitching’ grammar. As was discussed in the preceding chapter, several definitions of the matrix (or host, or base, etc.) language have been proposed. Apart from the sociolinguistic and discourse-related criteria, which I deem inappropriate for the identification of morpho-syntactic structures (cf. p. 38), all criteria refer to the ML as structuring the clause that contains morphemes from two or more languages. Provisionally I define the ML as ‘the language to which a grammatical structure (clause or clause constituent) containing morphemes from two or more languages can be attributed’.

Languages in contact with each other are seldom equal in status. In many cases one of the languages can be characterised as the bilingual community’s ‘own’ language, whereas the other language in use is imposed by an economically and/or culturally dominant speech community. For this broad generalisation I use the terms Community Language and Superimposed Language, which will be expanded upon in Chapter 11. The unequal status of the languages involved is reflected in the patterns

of intrasentential CS. However, this does not imply that we can attribute the morpho-syntactic frame of a mixed utterance to the grammar of a particular language on the basis of sociolinguistic considerations without examining this mixed utterance itself. A glance at the Moroccan Arabic/French examples in (1)-(9) hereafter and (48)-(49) in section 2.2 below reveals that either language may in turns provide the syntactic frame for a mixed sentence. Note that all these examples come from Slaoui's corpus of table conversations amongst her family members, and hence involve the same respondents in a rather constant conversational setting. There are correlations between the social status of the ML and the types of insertions from the other language that occur, but these can only be investigated if the definition of the ML is logically independent from sociolinguistic factors. For this reason the ML will be determined solely on the basis of the morpho-syntactic properties of the matrix structure itself, where this structure can be a clause or a constituent part of a clause. Subsequently the categories of embedded material can be examined and associated with sociolinguistic characteristics of the ML.

2.1.1 Why the concept of matrix language?

The concept of matrix language is in the first place a useful tool for the description of codeswitching data and the detection of regularities in these data. The following examples will demonstrate the advantages of the matrix language concept.

The Moroccan Arabic/French examples cited below can be described in a linear way as switches between French and Arabic elements. In terms of word categories, we make the observation that it is possible to switch between a French noun and a predicate adjective (or a \emptyset copula) as in (1), and between a French noun and an Arabic verb (2) or preposition (9). There can also be a switch between an Arabic verb and the French definite article, as in (3), (4) and (5). In addition, there can be a switch between the Moroccan Arabic demonstrative *hadik* or the indefinite article *wahed* and the French definite article, as in the last two examples.

- (1) *les restaurants mehlul-in hetta l l-wahd-a d l-lil*
 DEF-PL restaurant open-PL until to DEF-one-F of DEF-night
 "The restaurants are open until one in the morning."
 MA/French (Slaoui, 1986, Annexe I:6)
- (2) *la direction gal-et-l-hūm šuf-u mša l-ʔustad dyal-kūm*
 DEF-F management say-3F-to-3PL look-IMP-PL with DEF-lecturer of-2PL
 "The management said to them: settle [this] with your lecturer."
 MA/French (Slaoui, 1986, Annexe VIII:5)
- (3) *melli ta-ye-t-heyed l'email, ka-ye-t-šerrā-w*
 when ASP-3-MP-take-away DEF-enamel ASP-3-MP-be-naked-MP-PL
 "When the enamel gets loose, they [the teeth] become bare."
 MA/French (Slaoui, 1986, Annexe V:10)

- (4) gal-l-ek bġa y-dir *l'implantation*
 say-to-2SG want 3-do DEF-implantation
 "He said he wanted to have a [hair] implantation."
 MA/French (Slaoui, 1986, Annexe V:15)
- (5) ta-n-fiġ-u *le matin*, ka-n-šerġ-u atay nigru
 ASP-1-wake-up-PL DEF-M morning ASP-1-drink-PL tea black
 "We wake up in the morning and drink black tea."
 MA/French (Slaoui, 1986, Annexe II:7)
- (6) ħit ka-y-ġewwt-u *ħla le prof?*
 because ASP-3-shout-PL at DEF-M lecturer
 "Because they were shouting at the lecturer?"
 MA/French (Slaoui, 1986, Annexe VIII:12)
- (7) haduk lli ma *ħend-ħūm-š la pigmentation*
 DEM-PL REL NEG with-3PL-NEG DEF-F pigmentation
 "Those who don't have pigmentation." MA/French (Slaoui, 1986: 29)
- (8) gal-et-l-i hadik *la fille (...)*
 say-3F-to-1SG DEM-F DEF-F girl
 "That girl said to me (...)" MA/French (Slaoui, 1986, Annexe V:15)
- (9) dayr-in waħed *l'article ħla l'épidémio*
 do-PART-PL INDEF DEF-article on DEF-epidemiology
 "They did an article on epidemiology."
 MA/French (Slaoui, 1986, Annexe VI:6)

Alternatively, the data can be described in terms of grammatical functions, for instance, there can be a switch between a French Subject noun and an Arabic adjectival predicate (1) or verb (2). There can also be a switch between an Arabic verb and the French definite article in Subject position (3) or Object position (4), or in adverbial adjunct (5). In addition, switches can occur between an Arabic preposition and a French definite article in complement position, as in (6).

I will refrain from citing all switch points in these examples in every possible classification. Clearly, this is a cumbersome manner of description, since further examples will add still more possible juxtapositions while there are many types of classification. Moreover, an enumeration of possible juxtapositions is not an attractive way of presenting a grammatical description even if many studies on codeswitching have followed this method to a greater or lesser extent.¹ A systematic investigation

¹ This is probably a consequence of the initial impression that speakers actually switch from using one language to using another, like a farmer switches from stock breeding to arable farming, hence the term *codeswitching*.

of, for example, the functions of the noun phrase in this corpus or the distribution of the definite article in the text reveals that the linear approach is actually describing Moroccan Arabic grammar. Or, formulated differently, many aspects of MA/French codeswitching can be described by simply referring to MA grammar. The matrix language approach makes this explicit. With respect to the examples presented above we can make the generalisation that the distribution of French [definite article-N] parallels the distribution of Arabic [definite article-N] in Moroccan Arabic. Hence the description of the codeswitching in (1)-(9) can be summarized as follows:

- (10) French [*{le, la, l', les}* N] can be inserted in MA sentences wherever MA has a slot for [*I-N*].

Not only does this approach provide a simple and adequate description of the distribution of the juxtapositions of Moroccan Arabic and French in the above miniature corpus, it also explains why the distribution is the way it is.²

If the presence of the French definite article depended on French grammatical rules the French articles in (7), (8) and (9) would be out of place, whereas these articles are perfectly interpretable within Moroccan Arabic grammar. As for (7), MA uses the definite prefix *l-* to mark an undetermined quantity of non-count nouns, like *pigmentation* “pigmentation” in this example (Caubet 1993, I:186, II:274-6). In a translation equivalent of (7), French would use the ‘partitive article’ *de*: “*Ceux qui n’ont pas de pigmentation*” (Slaoui, 1986: 29). See also Slaoui (1986: 29-30) for a discussion and for similar examples with collective nouns. As for (8) and (9), Moroccan Arabic, but not French, uses the definite prefix after the indefinite article *waḥed* and after demonstratives.

Clearly this kind of description is not as theoretically neutral as the linear approach. The concept of insertion adds a level of explanation by assuming a fundamentally distinct role for each language, and by attributing the bulk of the attested regularities to the grammar of one of them, the matrix language. While the insertion approach has these distinct advantages, it also charges us with the task of formulating consistent criteria for the identification of the ML. Such criteria are essential when we are confronted with less clear-cut examples of mixed sentences. In addition, an inventory and classification of the possible insertions from the embedded language are required. Of course, such an inventory will be very small compared to an inventory of juxtapositions, as one type of insertion summarizes a large number of linear switch points.

Note that even if the matrix language model should fail to explain a particular phenomenon, it would still be a useful notion. Suppose that with respect to the example discussed above it turns out that the distribution of French [definite article N]

² The facts are somewhat more complicated, since the French masculine article *le* is usually replaced by the Moroccan Arabic definite prefix *l-*. See Wernitz (1993: 188) and also Boumans & Caubet (forthcoming), on Algerian Arabic/French.

in Moroccan Arabic sentences is not totally identical to the distribution of Arabic [definite article N], for instance because the French alternative is found to occur more often in pre-verbal position. In this case it would still be more convenient to state this finding as an amendment to the rule formulated in (10), than to describe all the occurrence patterns in terms of possible switch (juxtaposition) sites.

2.1.2 Identification of the matrix language

The point of departure is that the matrix language is a device to describe grammatical structures containing morphemes from more than one language. In this chapter we will be concerned with two main types of matrix structure, namely the finite clause and clause constituents headed by one of the major word categories Noun, Verb, Adjective, Adverb and Preposition. I will use the term matrix structure to refer to both types at the same time. The notion of ML attributes the morphosyntactic properties of the matrix structure to the grammar of one of the contributing languages rather than to both languages. The embedded language only contributes embedded elements, that is, it does not participate in creating the matrix structure. This chapter is only concerned with the identification of the ML in morphological and syntactic structures on the level of the clause, although in the next chapter some supra-clausal structures will be considered as well.

Since the ML is supposed to be responsible for the syntactic and morphological structure at hand, the grammatical structure of the mixed sequence inversely will identify the ML. The identification of the ML therefore depends on the interpretation of grammatical structures. As the ML will be identified in this manner it would be circular to predict that the ML provides the morphosyntactic structure. The only claim I make is that a ML can be consistently identified for all mixed structures.

Generalisation over codeswitching data, in the sense of reducing the number of patterns and explanations, is the most important device for the identification of the ML and embedded material. This means that whenever there are alternative ways to analyse a particular case the analysis that is most consistent with the rest of the data will be assumed. This approach will be developed and exemplified in section 2 below.

2.1.2.1 The constituent level

Let us start with an ordinary case of CONTENT MORPHEME INSERTION in (11).

- (11) t-ħafed ʔla l-*cultuur* dyal-ek
 2-preserve on DEF-culture of-2SG
 “You’ll preserve your culture.” MA/Dutch (Hocine)

The word order and the function morphemes can all be attributed to MA. The only unequivocally Dutch element is the noun *cultuur*. The analytic marking of possession *dyal-ek* is usual in MA, whereas the Dutch equivalent of this sentence would use a

pronominal possessive pronoun: *jouw cultuur* “your culture”.³ Likewise, in (12) not only are the function morphemes in Wolof, but the morpheme order N-Determiner must also be ascribed to this language rather than French. (The status of the French conjunction *ou bien*, which can scarcely be classified as a content morpheme, is the topic of Chapter 3, section 2.3.)

- (12) *cours bi moo intéressant ou bien ñi di def cours ñoo intéressant?*
 course DEF FOC interesting or those HAB do course FOC interesting
 “Is it the course that is interesting or those who take the course who are interesting?” Wolof/French (Meechan & Poplack, 1995: 182)

Similar examples can be found for content morphemes and constituents other than nouns and NPs. For instance, the English word *drive* in the following example can be analysed as part of a Cajun French VP *drive mon char*.

- (13) J’ai arrêté de *drive* mon char.
 “I stopped driving my car.” Cajun French/English (Brown, 1986: 401)

For the constituent level the ML can be defined as follows:

- (14) On the constituent level, the ML is the language to which the internal structure of the constituent as expressed by the distribution of all morphemes within the constituent can be attributed. The distribution of a morpheme concerns both its occurrence and its order relative to other morphemes that make up the constituent.

Hence the ML determines which function morphemes should or should not surface, and is responsible for the relative order of the function and content morphemes that make up the constituent. Note that it is the distribution of function morphemes, rather than the function morphemes themselves, that identifies the ML. For this reason, the absence of a function morpheme that is required according to ML grammar constitutes a counter-example to the MSA, see section 4.1.1 below. In addition, the above definition of the ML on the constituent level does not categorically exclude the insertion of function morphemes. This point will be dealt with presently.

Note that in itself, the idea of content morpheme insertion is not controversial. The definition in (14) and the discussion of examples (11)-(13) above are presented to demonstrate that even in apparently obvious cases the ML is identified by its role in structuring the mixed constituent rather than by the number of morphemes in a stretch of discourse or some un-specified intuitive criterion. The criterion in (14) is

³ An analytic genitive with pronominal possessor occurs with demonstratives in Dutch, but is as marked as its English equivalent, e.g. *die cultuur van jou* “that culture of yours”.

obviously reminiscent of Joshi's and Petersen's association of inflections and function words with the matrix language (Ch. 1, p. 34) and of Myers-Scotton's Morpheme-Order and System-Morpheme Principles (Ch. 1, p. 36). In accordance with its predecessors the Monolingual Structure Approach envisages content morphemes rather than function morphemes as embedded elements. However, contrary to these models the MSA does not a priori exclude the insertion of function morphemes in mixed constituents.

With respect to function morpheme insertion we can distinguish between two situations. One occurs when a mixed constituent contains a function morpheme from one language while the other morphemes and the distribution of all morphemes point to another language as the ML for that constituent. The definition of the ML in (14) does not exclude this situation but CS data provide little evidence for the occurrence of certain types of singly embedded function morphemes such as determiners, prepositions, case and agreement markers and inflectional morphology in general. This is not a premiss of the MSA. Quite the contrary, the MSA allows us to infer this generalisation from the data since it does not a priori exclude function morpheme insertion. Once we have made this observation in a principled rather than intuitive manner, we can start considering a possible explanation.⁴ Moreover, certain morphemes that might be classified as function morphemes do occur as singly embedded morphemes. As an example, see the discussion of time adverbs in the preceding chapter (p. 40).

The other situation occurs when complex word forms are inserted that consist of an EL content morpheme and one or more EL affixes or clitics. The insertion of derived and inflected content words is quite common, and subject to regularities. Recall the Moroccan Arabic/French examples above that were used to argue for a matrix language approach. The NPs *la pigmentation* in (7) and *hadik la fille* in (8), for instance, contain the etymologically French function morpheme *la*. However it was shown that the internal structure of the NPs is MA, and only MA as the ML can account for the occurrence of the French definite article in these examples. This is in accordance with the definition of the ML in (14) which states that the ML governs the distribution of function morphemes. A fairly common example of embedded grammatical morphology that is traditionally classified as inflection is the plural marking of nouns. In (15) we see the English plural suffix *-s* on the embedded noun *steak*. The inserted plural noun triggers (masculine) plural agreement in its modifiers in the Spanish NP (*unos* and *sabrosos*). In (16) the Dutch plural noun in the left-

⁴ Backus (1996b: 121-31) for instance, proposes an explanation in terms of semantic transparency and syntactic autonomy, two properties he subsumes under the cover term of 'specificity' in his Specificity Continuum. According to Backus, function morphemes like articles and inflectional morphology are not inserted because they have schematic meanings and their usage depends on the occurrence of other elements, in other words, because they are low on the Specificity Continuum.

dislocated Arabic NP *had ambtenaren* agrees in number with the Object suffix *-h ũm* that refers to it.

- (15) daban unos *steak-s* tan sabroso-s
 they·gave INDEF·M·PL steak-PL so tasty·M-PL
 “They served some steaks so tasty.” Spanish/English (Pfaff, 1979: 306)

- (16) had *ambtenar-en* ʔiden ka-ye-ŝti-w-hũm
 DEM civil·servant-PL thus ASP-3-give-PL-3PL
 “So these civil servants, they give them ..” MA/Dutch (Zineb)

The agreement phenomena suggest that the embedded plural suffixes (English *-s* and Dutch *-en*) mark the ML (Spanish, Moroccan Arabic) grammatical category ‘plural noun’. In other words, these EL plural nouns participate as ML plural nouns in the ML nominal paradigm. An instance of an EL inflected verbal form is presented below in (17).

- (17) sie sinn *condamn-és* worre
 they are convict-PASTPARTICIPLE become·PASTPARTICIPLE(PASSIVE)
 “They were convicted.”
 Alsatian German/French (Gardner-Chloros, 1991: 131)

The French word *condamn-és* can be considered here as an inserted element in the Alsatian German verbal constituent *condamn-és worre*. The French past participle marked by final *-é* [e] is used in a passive construction with the Alsatian German passive auxiliary as if it were an Alsatian German past participle. Hence the distribution of the EL verbal suffix *-é* is in accordance with ML grammar.

At this point a remark is in order concerning the terminology in (14) and in the remainder of this study. What is envisaged by the terms ‘content word’, ‘constituent’, ‘content morpheme’ and ‘function morpheme’? A primary delineation of these terms is given here; classificatory problems will be taken up in section 2. Further, note that the terms ‘insertion’ and ‘embedding’, ‘inserted’ and ‘embedded’ (etc.) will be used as synonyms.

Content morphemes or content words

One of the major tasks of the MSA and of matrix language approaches generally is the classification of EL material that is inserted in mixed constituents. Is it content MORPHEMES or rather content WORDS that are embedded? The examples just discussed show that the idea of content morpheme insertion as advanced in Myers-Scotton’s MLF model is too restricted to describe the attested insertion patterns. Furthermore, embedded compound words cannot be explained by content morpheme insertion alone, unless the ML and EL share the same structures of compound words. ‘Content word’ instead of ‘content morpheme’ is a broader term that covers the attested derived and inflected forms as well as EL compounds. However, ‘content word’ is a rather

indiscriminate expression introducing its own problems of demarcation. This is an even more serious drawback of the term ‘EL island’ or even ‘internal EL island’ used in the MLF model since these terms can refer to any combination of two or more EL elements which are framed by a larger ML constituent. Both the terms ‘EL word’ and ‘EL island’ obscure the fact that insertion of EL inflected content morphemes is both highly regular and restricted to some categories in certain data sets. The embedding of EL plural nouns, for instance, is fairly common while the insertion of inflected verb forms or case-marked nouns is rare. Secondly, the embedding of EL plural nouns seems to correlate with properties of the matrix language; EL plural nouns are less common when the ML is an agglutinative language like Turkish, Finnish or Swahili (cf. section 4.1.1 below). In order to describe and explain the regularities concerned with the embedding of complex word forms I will adopt the following approach: the insertion of content morphemes is assumed to be relatively unrestricted, and the insertion of particular inflected words forms is expressly mentioned.⁵ In addition I will use ‘embedded content word’ as a cover term comprising both single morpheme and complex word forms in contexts where this distinction is not essential, while recognising at the same time that the insertion of inflected word forms is highly constrained.

Constituent (I)

Constituents are units in the analysis of sentences as hierarchical structures. Constituent classes, such as DET, N, NP, PP etc., are identified primarily by distributional facts. Secondary considerations include semantic coherence, prosodic evidence, facts about the distribution of pronouns, and conjunction (Jacobson, 1995). No criterion for constituency is entirely unproblematic, however. Alternative analyses of syntactic structures are possible and one can often opt for a more refined classification of constituents that takes smaller distributional differences into account.

Obviously, it is beyond the scope of this study to discuss what constituent types could or should be distinguished for any language. I will mostly be discussing the relatively well established constituent types NP, PP, AdvP, AdjP and VP. VP includes at least the lexical verb and its complements. The constituent that includes Tense and Aspect as inflectional categories will be called the finite clause (to be dealt with below). In other cases I will point to arguments for constituency. In relation to content morpheme insertion, only constituents with lexical (content word) heads are relevant.

The constituents at the bottom of the hierarchical tree, also called ultimate constituents, consist of just one morpheme. I will not be referring to ultimate constituents as constituents unless the discussion requires this. Obviously, constituents that serve as matrices will not be ultimate constituents.

⁵ The conditions that enable the insertion of complex word forms are the topic of Boumans (1995b; forthcoming).

Content morpheme and function morpheme

Content and function morphemes are complementary categories, so that the definition of one will define the other.⁶ With respect to the content/function morpheme dichotomy in codeswitching, Muysken (1995: 183) points out that there is no single valid criterion for an overall distinction. Rather, he presents four criteria. The first one is the open/closed class distinction (cf. Joshi, 1985). Typical open class morphemes are nouns and verbs, while pronouns generally belong to the closed class. The closed/open class distinction is not always a very sharp one, however. For instance, many languages have a limited number of conjunctions and adpositions, but it is not uncommon that elements are added to these categories. There is no distinction of open and closed class categories that applies to all languages: adjectives, for example, form an open class in many languages, but a small, closed class in others.

The second criterion is “whether a given closed class is paradigmatically organised, i.e., whether the elements are defined in opposition to each other (present vs past, singular vs plural, definite vs indefinite etc.)”. Personal pronouns and tense systems in particular tend to be tightly organised paradigmatically.

The third criterion is the morpheme’s role in structuring the clause. Muysken mentions subordinating conjunctions, tense and agreement markers as playing a central role in the clause.

Muysken’s fourth criterion is whether a morpheme is bound or free. In many, but not all, languages bound morphemes are function morphemes. But in so-called polysynthetic languages, for instance, we find morphemes that are content morphemes according to other criteria but only occur as bound morphemes. Likewise, in so-called isolating languages morphemes that are function morphemes by all other criteria will be free morphemes.

In sum, the prototypical function morpheme belongs to a closed class, is a paradigmatically organised bound morpheme, and plays a role in structuring the clause. However, there are many examples of morphemes that range under content morphemes according to one criterion, and under function morphemes according to another criterion. Subordinating conjunctions, for instance, play a role in structuring the clause, but they are not paradigmatically organised, nor are they necessarily bound morphemes. While there is no confusion with regard to the prototypical classes of content and function morphemes, remaining intricate cases cannot be categorically assigned to either class by any single criterion, nor by a set of associated criteria. The problematic cases include pronouns, some adpositions, aspectual and time adverbs and conjunctions. These categories prove problematic in the analysis of

⁶ While the terminology used here has clear affinity with Myers-Scotton’s terms ‘content morpheme’ and ‘system morpheme’, remember that the latter terms have very specific meanings in Myers-Scotton’s MLF model (see Chapter 1).

codeswitching, since their ‘codeswitching behaviour’ does not systematically match that of either content or function morphemes.

Personal pronouns, for example, are not commonly inserted elements and in this respect they resemble function morphemes. But English first and second person pronouns are pervasive in bilingual varieties of Malay (Ozóg, 1987) Indonesian and Thai (Foley, 1986: 210). These facts suggest that the insertion of a morpheme does not depend just on its status as content or function morpheme in the source language, but also on properties of the corresponding category in the matrix language.⁷ The discussion of EL pronouns in general and in Malay/English in particular will be continued in 2.2 below. See also Chapter 1, p. 40, 41 and Jake (1994) for a classification of pronouns as either system or content morphemes within Myers-Scotton’s MLF model.

Remember that the content/function morpheme distinction is not a key concept in the Monolingual Structure Approach. The MSA shows that single function morphemes are not usually embedded, however this is an observation (see section 2.1 below), not a premise of the model. The definition of the ML on the constituent level circumvents the intricate distinction between function and content morphemes by referring to the distribution of all morphemes that make up the constituent, regardless of their source language. Note that, since the ML on the constituent level is inferred from its internal language-specific properties, the association between the ML and the function morphemes is nonetheless preserved. After all, it is the distribution of function morphemes rather than content morphemes that is language-specific and designates the ML. A second advantage of this approach is that it allows us to investigate exactly what kinds of morphemes do occur as embedded elements, and under which circumstances.

2.1.2.2 The finite clause level

Now consider (18) below.

- (18) *un risque de la condensation* heb je
 INDEF risk of DEF.F condensation have you
 “You do have a condensation risk.”
 Brussels Dutch/French (Treffers-Daller, 1994: 220)

⁷ With respect to Thai and Indonesian, Foley (1986: 210) reports that English *I* and *you* function as neutral pronouns next to the indigenous forms that indicate social status. Bert Tahitu (personal communication) informed me that in some idiolects of Moluccan Malay spoken in the Netherlands the Dutch pronoun *ik* ‘I’ is used for the same reasons. The use of EL pronouns may thus be motivated by ‘lexical gaps’ that result from changes in the social organisation.

In this sentence the French function morphemes outnumber the Dutch ones. The French function morphemes are, however, concentrated in one NP constituent. Moreover, the constituent order is recognisably Dutch. The so-called ‘verb second’ rule places the finite verb after the first constituent, in this case the topicalised complement of *heb* ‘(you) have’. Brussels French also allows for Object topicalisation but does not have the ‘verb second’ rule (Treffers-Daller, 1994: 220).

Conversely, the constituent order in (19) suggests French as the matrix language, since in Dutch non-finite verbs are clause-final and can only be followed by (heavy) prepositional phrases. In (20) the order Verb-Subject in a declarative sentence is in accordance with Moroccan Arabic, not Dutch syntax, and in (21) the order Verb-Object is English whereas Farsi has Object-Verb. Finally, in (22) we find the Turkish constituent *baba-m1z, anne-miz* embedded in a Dutch sentence. Here too the constituent order is decidedly Dutch, since Turkish is verb-final. Furthermore, the Turkish NP is the complement of a Dutch preposition while Turkish has only postpositions and post-nominal case marking, cf. *baba-m1z, anne-miz gibi* ‘like our parents’.

- (19) *je dois, je dois gliss-er* daan vinger hier
 I must I must slide-INF that finger here
 ‘I have to press here with my finger.’
 Brussels Dutch/French (Treffers-Daller, 1994: 220)
- (20) *Ťta-ha-ni de buurman*
 give-3F-1SG DEF neighbour
 ‘The neighbour gave it to me.’ MA/Dutch (Hocine)
- (21) *you’ll buy xune-ye jaedid*
 you’ll buy house-POSS new
 ‘You’ll buy a new house.’ Farsi/English (Mahootian, 1993: 152)
- (22) *we zijn gewoon het-zelfde als baba-m1z, anne-miz*
 we are simply the-same as father-1PL mother-1PL
 ‘We’re just the same as our parents.’
 Turkish/Dutch (Backus & Boumans, 1996: 150)

Examples like (18)-(22) suggest the recognition of NP INSERTION as a codeswitching mechanism. The identification of constituent insertion has two advantages: firstly, it explains the constituent order in examples such as (18) (22), secondly, it licenses the occurrence of embedded language function morphemes on the condition that they are part of an EL constituent.

Not only NPs are inserted; other types of embedded constituents can also be identified. The next two examples show that CONSTITUENT INSERTION as a CS mechanism is a warranted generalisation. Example (23) is parallel to (18) above: here too, the constituent order is recognisably Dutch, including the application of the ‘verb second’ rule. But this time, the topicalised constituent is a Moroccan Arabic PP.

- (23) mʃa-k *ben ik mezelf*
 with-2SG am I myself
 “With you I’m being myself.” MA/Dutch (Samir)

The following example shows VP insertion:

- (24) ja będę siedzieć i *eat my meal*
 I be sit·INF and eat my meal
 “I’m going to sit and eat my meal.” Polish/English (Ewing, 1984: 52)

Here the Polish infinitive *siedzieć* “to sit” and *eat my meal* are conjoined by *i* “and”. The conjunction argues for the constituent status of *eat my meal*, and therefore for constituent insertion as a CS mechanism. This follows from the general assumption that constituents conjoin with constituents of the same category, together forming a constituent of the same category (Jacobson, 1995: 718). Hence in this example it is coordination rather than word order that supports constituent insertion (cf. Bautista, 1980: 64-5).

Constituent insertion is demarcated as follows:

- (25) The EL constituent is a constituent that is well-formed according to EL grammar, and has the distribution of an ML constituent in the ML matrix structure. The distribution of a constituent concerns both its mere occurrence and its order relative to the other parts with which it forms a higher-order constituent.

Despite the fact that in principle there can be evidence for many types of embedded constituents, in most data corpora it can be observed that constituent insertion in CS is restricted to a few constituent types, typically NP and PP. I assume these constituents to be embedded in a larger structure called FINITE CLAUSE.

Levelt (1989: 256-8) defines the distinction between basic and finite clauses (in English, at least) as follows. A basic clause contains one and only one main verb, tensed or otherwise; every finite clause contains one and only one finite verb. There is evidence that the basic clause is the primary planning unit in speech production and on this level grammatical functions are assigned to arguments. But the finite clause “concerns the ordering aspect of grammatical encoding; it reflects which categorial procedures do the word and phrase ordering of the retrieved functional information” (Levelt 1989: 256). In the following, I take Levelt’s finite clause as the next relevant matrix structure above the constituent level. It may prove necessary, however, to recognise other types of matrix clauses as well, e.g. imperative clauses.

Constituent (II)

In section 1.2.1 above the constituent was discussed as a matrix structure in which content (and possibly function) morphemes are embedded. Now that we are examining constituents as EL elements in a matrix structure defined as finite clause, we will consider the notion of constituent once again.

While the finite clause is relatively well defined, the delimitation of ‘constituent’ is left open-ended except for the requirement that the EL constituent and the ML constituent whose position it occupies in the ML frame be established constituent types in the respective languages. That is, language-specific information should provide the criteria to identify constituent types in the various languages concerned. Remember that distributional facts are central for the establishment of constituency. Various kinds of constituents are postulated by various syntactic theories, and in principle each constituent type may also occur as an EL constituent. The matrix language concept serves to investigate what types of EL constituents do occur. Indeed, codeswitching data provide additional evidence for constituency.

The constituent viewed as a unit in the hierarchical structure analysis of sentences may itself consist of several smaller constituents. A finite clause can itself be a constituent within another clause as is the case with complement clauses or relative clauses. So, in theory, there can be an infinite number of hierarchically ordered matrix constituents, and a different ML may be identified for each level. In practice, constituent insertion is usually restricted to a few constituent types in each data corpus. Recall that on the lowest level of analysis each morpheme is formally an ultimate constituent. I will avoid referring to ultimate constituents as EL constituents except in those cases where distributional criteria place them on a par with more complex constituents. For example, a single adverb (e.g. *often*) often has the same distribution as more complex adverbial constituents (*very often*). In the same manner distributional criteria identify pronouns as a kind of NP. This will be discussed further in section 2.2 below.

Definition of the ML

If constituent insertion is a separate CS mechanism besides content morpheme insertion, the ML must be defined independently for mixed constituents and mixed finite clauses. Note that the distribution of function morphemes cannot be simply used to define the ML on the finite clause level. After all, function morphemes are part of constituents that are themselves potentially EL constituents. A quantitative criterion such as ‘the language of the majority of the morphemes’ will not work either. Remember that in (18) there are more French function words than Dutch, while the constituent order points to Dutch as the ML (cf. the discussion of the MLF model in Chapter 1, p. 37). ML in mixed constituents is not an effective criterion for the ML on the finite clause level either. In the first place because there may not be any mixed constituent, as in the above examples, and secondly, because it is possible that the ML on the constituent level is different from the ML on the sentence level as I will demonstrate shortly (see also Chapter 1, p. 47). In principle and in analogy with

the definition of the ML at the constituent level the ML on the finite clause level can be inferred from the internal structure of the clause, viz. the selection and ordering of clausal constituents.

Based on this reasoning, I have so far relied on constituent order to argue for the concept of constituent insertion, equating the language of the constituent order with the ML in the discussion of examples (18)-(23). Nishimura (1986) actually proposes word order in the sense of constituent order as the criterion by which to define the ML and this criterion is implicit in Sridhar & Sridhar (1980). The problem of this criterion is that it only helps to pinpoint the ML in mixed clauses exhibiting a word order that is precluded in the syntax of one of the contributing languages. There is no reason to assume that constituent insertion is limited to cases of divergent word order. Consider another example from Treffers-Daller's (1994) corpus. Here, unlike in the Brussels Dutch/French clauses cited above, the constituent order does not point to either language as the ML since Verb-Object is the normal word order in both French and Dutch.

- (26) daar zetten ze euh *des* *barrières*
 there put they er INDEF·PL barriers
 "There they put up eh barriers."
 Brussels Dutch/French (Treffers-Daller, 1994: 204)

Concerning (26) and similar examples, one could in effect argue that there is no insertion and that each constituent is placed according to the rules of the language from which its morphemes derive. However this reasoning cannot be maintained in those cases where there is a word order conflict like in the other Dutch/French examples (18) and (19). Conversely, all three Brussels Dutch/French examples can be analysed as cases of NP insertion. This has the advantage of reducing 'constituent insertion' and 'juxtaposition in case of equivalent word order' to a single code-switching mechanism, i.e., constituent insertion (see also the evaluation of the Equivalence Constraint in Chapter 1 p. 14 ff.).

Of course the matter is not limited to Brussels Dutch/French. The order Verb-Complement, for instance, is common to many languages and embedded Object NPs may be recognised in other data corpora as well. Some further examples in which both ML and EL share the word order Verb-Object are given below:

- (27) (no creían en Jesús) *and then he sent* este hombre
 they didn't believe in Jesus this man
 "They didn't believe in Jesus, and then he sent this man .."
 Spanish/English (Lipski, 1978: 256)
- (28) ġadi ne-fhem *de* *hoofdlijn-en*, (*zeg maar, die begrijp ik wel*)
 FUT 1-understand DEF outline-PL say, those understand I AFFIRM
 "I'll understand the outlines, say, I do understand those."
 Moroccan Arabic/Dutch (Abdellah)

- (29) a neus bet *une fracture du* *crâne*
 have·3SG PRES have·PASTPARTICIPLE INDEF fracture of·DEF skull
 “She has had a fractured skull.” Breton/French (Timm, 1994: 109)

In order to identify the Object NPs in (26)-(29) as embedded constituents and thus bring them under one label together with the earlier examples (19)-(24) a definition of the ML on the finite clause level is required that does not depend on word order. In the case of the finite clause there is fortunately a suitable independent criterion: the verbal inflection, or perhaps more precisely inflection for tense, is probably the best indicator of the ML. With respect to all of the examples reproduced in (19)-(29) it turns out that there is a strong correlation between the language to which basic word order must be attributed and the finite verb. In other words, the same language that provides the inflection of the tensed verb also organises the relative order of the verb and its arguments. Since there is more variation in verbal inflection systems than in constituent order the finite verb is the most constant criterion.

Indeed, a number of scholars define the matrix language as the language of the verb (Wentz, 1977; Pandit, 1986; see p. 34 ff in Chapter 1). Slaoui (1986), writing on Moroccan Arabic/French also identifies a mixed sentence as ‘close to one language’ if the predicate appears in that language. From her examples it appears that by predicate, Slaoui actually means the verb (1986: 17-18). Klavans (1985) provides the required precision in identifying the inflection bearing element of the verb as indicative of the ML. Klavans’ conclusion receives support from Treffers-Daller (1994: 204). In the Monolingual Structure Approach, Klavans’ (1985) definition of what she calls base language is adopted for the ML on the finite clause level:

- (30) The Matrix Language (ML) on sentence level is the language of the inflection bearing element of the tensed verb.

At this point I can only infer that the verbal inflection and the constituent order tend to be attributable to the same language. It is not entirely clear why this is the case, although the relationship between verbal inflection and basic word order has been observed before in linguistic theory. So Treffers-Daller defines the language of the finite verb as the base language of the sentence “because the sentence is defined as the maximal projection of inflection (I) in modern linguistic theory (IP)” (1994:204). And, as Moyer (1995: 194-5) points out, “Ouhalla’s (1991) functional categories proposal does provide theoretical support for Klavans’ view of INFL as an important category in code-switching. According to Ouhalla the ordering of TENSE and AGREEMENT categories with respect to each other are shown to be ultimately responsible for word order variation across languages”.

It must be emphasised that it is the verbal inflection that correlates with the word order, and not the lexical verb itself. Non-tensed verbal forms can be inserted in an ML frame in a number of ways (cf. Boumans, 1995a; 1996). In codeswitching involving two European languages of the Indo-European stock it is quite common

to find non-finite verbal forms such as participles and infinitives inserted as EL constituents, see number (24) above, or as ‘inflected content morphemes’, see (17). Apart from this, EL verb stems or infinitival forms are often combined with ML inflectional morphemes to make finite verbs. Either the ML inflection is attached to the EL verb (31), or ML inflection is expressed on a kind of dummy verb in the so-called *do*-construction, as in (32).

- (31) ní *dispute*-alfaidh mé leat
 NEG dispute-FUT I with·you
 “I won’t dispute with you.” Irish/English (Stenson, 1990: 180)

- (32) ka-t-dir mŋa-hŋm *voetball-en*?
 ASP-2-do with-3PL play·soccer-INF
 “Do you play soccer with them?” MA/Dutch (Samir)

Note that in the Irish/English example, the order Verb-Subject-Complement coincides with the language of the verbal inflection (Irish), not the language of the lexical verb.

2.1.2.3 Layered insertion

In section 1.2.1 it was shown that (inflected) content morphemes from one language are inserted in constituents that as a whole are associated with another language, the ML. In addition constituents from one language were shown to be inserted in finite clause frames that are associated with another language (section 1.2.2 above). Criteria for the identification of the ML on the constituent and on the finite clause level were presented in (14) and (30). Because an ML can be identified independently on both levels, the ML on the finite clause level is not necessarily the ML of every constituent forming a part of the clause. If content morphemes and constituents are inserted it must be at least theoretically possible to have an inserted content morpheme within an inserted constituent. Examples of this are actually quite common in some data sets. In the concluding section of Chapter 1 (p. 47) I have briefly touched on the concept of LAYERED INSERTION in relation to government approaches to CS. Here this concept will be worked out from the perspective of the MSA.

Bautista (1980) and Nishimura (1986) propose a layered insertion analysis for some of their Tagalog/English and Japanese/English data, respectively. The following example sentence from Nishimura will illustrate the point.

- (33) *I slept with her basement de*
 LOCATIVE
 “I slept with her in the basement.”
 Japanese/English (Nishimura, 1986: 130)

Nishimura analyses the entire clause *I slept with her basement de* as an English sentence structure. See the tree representation of (33) in Fig. 2.1 where *e* and *j* stand for English and Japanese respectively and crossing lines indicate insertions. In this

English sentence a Japanese postpositional constituent is embedded: *basement de*. This embedded PP forms itself a matrix structure in which an English noun is embedded.⁸

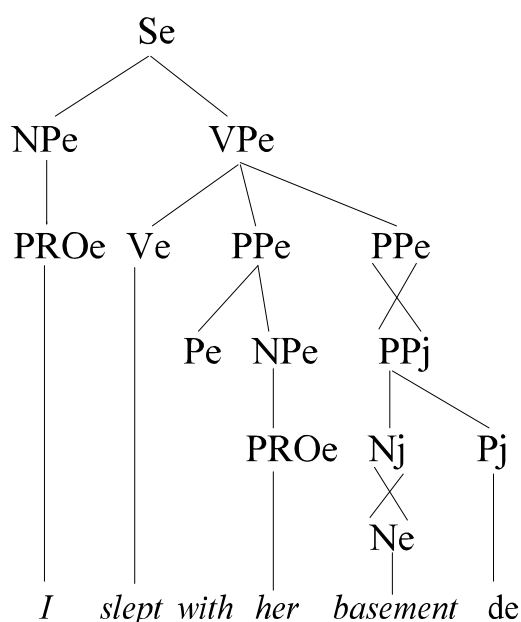


Fig. 2.1. Layered insertion.
(Nishimura, 1986: 131)

Nishimura's analysis is perfectly in line with the identification of the ML on separate levels as outlined above. The alternative, to consider just the Japanese postposition *de* as an insertion, is unattractive for two reasons. Firstly, it would necessitate the introduction of 'function morpheme insertion' as an additional insertion type, a type, moreover, for which there is little evidence in codeswitching data generally. Secondly, even if one accepts the possibility of opposition insertion for Japanese/English the absence of a determiner and the relative order of noun and adposition in *basement de* argues against the analysis of this constituent as an English PP.

Further evidence for layered insertion is found in Moroccan Arabic/French code switching on which there is much documentation (e.g. Abbassi, 1977; various articles by Bentahila & Davies since 1983; Slaoui, 1986; Nait M'barek & Sankoff, 1988; Lahlou, 1991; Wernitz, 1993). In all the available data, insertions of French [definite article N] are abundant, review examples (1)-(9) at the beginning of this chapter. Embedded Arabic noun phrases and prepositional phrases in French finite clauses are also attested, although these are less frequent. (34) and (35) provide examples of this:

⁸ In her critique of Nishimura (1986), Romaine (1995:145) seems to overlook the aspect of layered insertion.

- (34) *attention il ne faut pas chang-er t-tewṣil*
 caution 3SG·M NEG should NEG change-INF DEF-receipt
 “Be careful not to change the receipt.”
 MA/French (Bentahila & Davies, 1991: 383)
- (35) *ils prennent un fil de fer men berṛa*
 they take a string of iron from outside
 “They take an iron string from outside.”
 MA/French (Slaoui, 1986, Annexe 3:12)

The combination of (inflected) content morpheme insertion and constituent insertion yields cases of insertions within insertions:

- (36) *tu perds waḥed l-demi-heure*
 you lose INDEF DEF-half-hour
 “You lose half an hour.” MA/French (Bentahila & Davies, 1991: 383)
- (37) *c’est un pédé homosexuel parmi haduk les frères musulmans*
 he·is a faggot homosexual among DEM·PL DEF·PL brother muslim
 “He’s a faggot among those Muslim Brothers.”
 MA/French (Wernitz, 1993: 286)
- (38) *il a été convoqué f dak les premières convocations*
 he has been summoned in DEM DEF·PL first·F summoning
 “He was summoned in those first summonings.”
 MA/French (Bentahila & Davies, 1991: 383)

Employing the Monolingual Structure Approach, (36) is analysed as a French finite clause with an inserted Arabic NP (*waḥed l-demi-heure*) in which a French content word is inserted (*demi-heure*). (37) is a parallel example with the demonstrative *haduk*. Analogously, in (38) the inserted Arabic PP *fdak les premières convocations* constitutes a matrix on the constituent level that embeds the French inflected content morphemes *les premières convocations*.⁹

Notice how the layered insertion analysis makes use of the previously established categories of (inflected) content morpheme and constituent insertion, examples of which are to be found in the same data corpora. Again, the strings *waḥed l-*, *haduk*, and *fdak* in the above examples cannot be considered as Arabic insertions for a

⁹ An insertion such as this could be called a constituent (NP) insertion. It should be kept in mind, however, that the distribution of embedded French [definite article N] in Moroccan Arabic is clearly not that of Arabic NP constituents as a whole. In addition the insertion of [definite article N] sequences and of adjective-noun combinations is subject to fewer constraints than other EL NPs. We return to this in the next section.

number of reasons. The category ‘function morpheme insertion’ is not favoured in the Monolingual Structure Approach because it falls outside normal patterns in CS (see Bautista, 1980: 52, for a similar argument). But even if we assume the possibility of function morpheme insertion, we must consider in what structure these Arabic function morphemes are embedded. Recall that the sequences [indefinite article - definite article] in (36) and [demonstrative - definite article] in (37) and (38) are perfectly interpretable within Moroccan Arabic grammar, whereas they would be ungrammatical sequences in French NPs. By this process of reasoning the insertion of Arabic demonstratives and indefinite articles becomes even less plausible in these examples.

The layered insertion analysis offers a solution for seemingly problematic cases of function morpheme insertion in a variety of other language pairs. Mahootian & Santorini (1996), citing (39) and (40), rightly indicate that such cases cannot be explained within a model that identifies an ML at only one level.

- (39) *where are they, los language things?*

DEF·PL·M

“Where are they, the language things?”

Spanish/English (Poplack, 1981: 175)

- (40) *maen through taemam-e ina raeftaem*

I through all-POSS these I·went

“I went through all of this.” Farsi/English (Mahootian, 1993: 111)

- (41) *et ça était blanc comme de la neige et zo’n belle petite trémousse*

and that was white as INDEF snow and such·a nice little jumper

“And it was white as snow, and such a nice little jumper.”

Brussels Dutch/French (Treffers-Daller, 1994: 119)

Therefore I propose that *los language things* in (39) is a Spanish NP inserted in an English matrix clause; *through taemam-e ina* in (40) is an embedded English PP; and *zo’n belle petite trémousse* in (41) is a Dutch NP. Layered embedding sometimes results from repetition of a mixed constituent as Eliasson (1995: 51-2) convincingly demonstrates. He presents the following passage from a Maori/English narrative:

- (42) *That’s right, i te haere mātau ki Mātauri, ā, ka pāhi mai te .. te .. te .. te ..*
te hearse, e tū ana mātau i tō mea raka, i tō Hōhepa rā, nē, te pick up i a
Hōhepa. Ana, ka karanga atu, “Gee, what’s te hearse over there?” Nā, ka
karanga mai ētahi, “Ana, ko Haki, kei te whakahoki mai i Whakatāne.”

“That’s right, we were journeying to Matauri and the .. the .. the .. the .. the hearse went by, while we were assembled at what’s his name’s place, the home of Hohepa. Aye, we were picking Hohepa up. I called out “Gee, what’s

the hearse over there?” Then some people replied, “It is Haki, being taken home from Whakatane.” Maori/English (Eliasson, 1995: 51)

At the beginning of the second line the English noun *hearse* is inserted; together with the Maori definite singular article *te* it constitutes the mixed NP *te hearse* “the hearse”. So far there is nothing remarkable about this example. But then the NP *te hearse* is repeated in its entirety as a constituent of the English matrix clause *what’s te hearse over there?* in the third line. The ML of this clause is English of course, according to the finite verb criterion. At first sight the Maori article *te* seems to be the embedded element, but the repetition of *te hearse* supports the idea that this is really an embedded Maori NP containing an English noun. In this way we can account for what seems to be a counter-example to the generalisation that single function morphemes are not inserted.

The above examples of layered insertion constitute a further argument that the ML must be identified on (at least) two levels in the hierarchic syntactic structure. According to Mahootian & Santorini (1996: 476 n. 14) the subsequent levels of embedding in Nishimura’s 1986 article are inconsistent with Joshi’s (1985) conception of ML. Yet the idea of layered insertion is not really a substantial digression from the current matrix language models that, like Joshi’s, recognise constituent insertion. After all, the basic concept of matrix language entails that this language sets the morpho-syntactic frame. Although the concept of matrix language is only relevant concerning the discussion of codeswitching, an ML can in principle be attributed to the morpho-syntactic structure of monolingual and mixed constituents alike. As soon as one recognises constituent insertion one is in fact assuming that one language is the ML on sentence level while another language projects the grammatical frame of the embedded constituent. It is this assumption together with the rather uncontroversial notion of content morpheme insertion that leads to the proposed layered insertion analysis.

2.2 Classification of embedded elements

Now that the matrix structures have been identified, the next step is to investigate what kind of elements are embedded. For an efficient syntactic description of CS behaviour and for an attempt at theoretical explanation it is essential to form generalisations over individual cases of insertion and to classify these into the smallest possible number of categories. This subsection will attempt to clarify some aspects of this decision-making process.

Of course, the general aim of classification is to be economical and conservative. The classification seeks to a) limit the number of categories and b) avoid unwarranted generalisations. Because a) and b) are conflicting aims classification is a matter of balancing pros and cons. Within a particular data set, for instance, two small, specific categories of embedded material, labelled *x* and *y*, may be reduced to one large class

Q. This halves the number of insertion categories. But if class Q is usually understood to encompass the categories *x*, *y* and *z*, Q is an unwarranted generalisation. Another undesirable method of generalisation expands a general class to encompass categories it is not usually understood to comprise, as when categories *w*, *x*, *y* and *z* are packed together in class Q which traditionally contains only the latter three.

The data set for which the classification is to be valid is of crucial importance. A large data set such as ‘all published CS material of all language pairs’ will contain more specific categories of embedded material and, accordingly, wider generalisations can be made. (That is, all of categories *x*, *y* and *z* will probably occur somewhere in the data set, so that the generalisation Q is easily justified.) In classifying embedded material in a smaller data set, e.g. ‘the contributions of one speaker to the Nijmegen corpus of Moroccan Arabic/Dutch’, one has to be conservative with regard to this particular data set while at the same time keeping in line with the tendencies observed in the larger set of MA/Dutch data. In the descriptive part that will follow in Chapters 4 to 10, the MA/Dutch data as a whole are considered against the background of code-switching data in general. This adds a further consideration: a classification that is attractive for a particular small set of data may be unattractive in the context of a larger set. The classification process is exemplified below.

2.2.1 Content versus function morpheme insertion

Principles of generalisation lead to the inference that it is content morphemes that are embedded in mixed constituents rather than function morphemes. Consider the next example to see how this works:

- (43) *had ambtenar-en* ʔiden ka-ye-ʃti-w-hũm ..
 DEM civil·servant-PL thus ASP-3-give-PL-3PL
 “So these civil servants, they give them ..” MA/Dutch (Zineb)

The Dutch plural noun *ambtenaren* “civil servants” in (43) can be analysed as an embedded content morpheme in a Moroccan Arabic NP. This analysis is in line with the observation that content morphemes, and in particular nouns, are often inserted in CS generally. However, noun insertion alone does not explain the configuration of the NP *had ambtenaren*: the plural suffix *-en* is also Dutch, and the NP lacks the definite article prefix that obligatorily follows the demonstrative in Moroccan Arabic. The MA equivalent of this constituent would be *had l-muwaddaf-in* (DEM DEF-civil·servant-PL). Superficially, the mixed NP *had ambtenaren* has more in common with its Dutch translation *deze ambtenaren* (DEM civil·servant-PL).

For these reasons, another analysis may at first glance seem more attractive: *had ambtenaren* is a Dutch NP, and the Arabic demonstrative *had* is the inserted element.¹⁰ This Dutch NP is further embedded as a left-dislocated constituent in an

¹⁰ Utz Maas and Jonathan Owens suggested this possibility to me.

Arabic sentence; in the finite clause the Object suffix *-h ũm* refers to the foregrounded NP. This alternative analysis - a Dutch NP - explains the Dutch plural suffix as well as the missing MA article, since in Dutch, demonstratives cannot precede a definite article.

However, this second analysis loses its appeal in the light of a larger sample of Moroccan Arabic/Dutch data. EL plural markers and missing MA definite prefixes turn out to be characteristic of MA/Dutch codeswitching; see Chapter 5 section 1.3 on the former, and section 1.5 on the latter for a detailed examination. The absence of an obligatory function morpheme, as in the case of the missing definite article in (43) and (44) below, is a complication for the MSA that will be addressed below in section 4.1.1 below. This phenomenon cannot generally be explained by assuming a Dutch NP. In (44) for instance, both MA and Dutch grammar would assign a definite article to *mensa* because the student restaurant in this example is identifiable to all the interlocutors.

- (44) beʃd l-xetr-at ka-ne-lqa-h f mensa teht ʃend-na
 some DEF-time-PL ASP-1-find-3M in student.restaurant below at-1PL
 “Sometimes I find it [a periodical] in the student restaurant below.”
 MA/Dutch (Abdallah)

- (45) f d-din, d-din ntaʃ l-christen-en, l-marʔa naqʃa
 in DEF-religion DEF-religion of DEF-christian-PL DEF-woman inferior
 “In the religion, in the Christians’ religion, the woman is inferior.”
 MA/Dutch (Maryam)

Concerning embedded plural nouns, this is a common insertion type in many language pairs, as noted earlier (cf. Boumans, 1995b; forthcoming). Example (45) shows that the assumption of NP insertion does not satisfactorily explain the occurrence of Dutch plural suffixes on Dutch EL nouns in MA/Dutch. The NP *l-christenen* in this example is one of the relatively few cases where the MA definite prefix does attach to an EL noun. Here the analysis *l-christenen* as a Dutch NP helps to explain the Dutch plural suffix but creates a new problem as the prefix *l-* would be an embedded function morpheme in this NP.

Generalisation engenders further arguments as to why *had ambtenaren* in (43) is not a Dutch NP. Concerning codeswitching in general, analogous argumentations lead to the conclusion that insertion of lone determiners such as the MA demonstrative *had* in (43) if *had ambtenaren* were to be considered as a Dutch NP, is unusual. For these reasons the initial analysis of (43) is the most plausible one.

2.2.2 Single morpheme constituents

A single content or function morpheme (ultimate constituent) can by itself constitute a higher order constituent. Whenever this applies a choice must be made between

two possible classifications: is the ultimate or the higher order constituent the element that is embedded?

Adverbs

Single adverbs, for instance, may constitute a complete adverbial phrase. Thus in (46), Dutch *direct* can be classified either as an EL adverb or an EL adverbial phrase.

- (46) *nee, direct* *değil, yedi ay sonra git-ti-m*
 no immediately NEG seven month after go-PRET-1SG
 “No, not immediately, I went after seven months.”
 Turkish/Dutch (Backus, 1996b: 140)

In the case of adverb insertion, it is the Turkish grammatical rules that shape the AdvP; in the case of an embedded AdvP, the AdvP is the product of Dutch grammar. The question is which grammar regulates the distribution of function morphemes in the constituent *direct*, that is, the absence of any function morpheme (e.g. adverbial marker) in this case. Since both Turkish and Dutch have adverbial phrases that consist of a single adverb, the question cannot be answered on the basis of an isolated example. Investigation of a larger sample of Turkish/Dutch will reveal whether Dutch adverbs are embedded as content morphemes in more complex Turkish AdvPs, and whether more complex Dutch AdvPs are embedded as EL constituents. An example of the former would be the modification of Dutch *snel* “quickly” by a Turkish degree adverb as in *daha snel* “more quickly”; the whole would be classified as a Turkish adverbial constituent by virtue of criterion (14). The occurrence of Dutch *veel te snel* “far too quickly” in a Turkish finite clause would be an unequivocal example of ‘adverbial constituent insertion’. If both phenomena appear to be common, the exact classification of *direct* in (46) is inconsequential. However, if only adverb insertion is attested, example (46) furnishes insufficient evidence for the insertion of AdvPs. And vice versa, if AdvP insertion is a possible classification for all cases, there is no evidence for classifying adverbs as EL content morphemes in this corpus.

Pronouns

Free form pronouns are another case in point. In terms of distribution, free form pronouns are full NPs in many languages. Thus EL pronouns may be considered as either embedded content or function morphemes (ultimate constituents), or as embedded NPs. The ‘EL constituent’ analysis emphasizes the distributional properties of pronouns, while the ‘EL function/content morpheme’ analysis stresses their surface form as single morphemes. The earlier discussion of example (36) in Chapter 1, repeated here as (47), becomes relevant at this point.

- (47) (*you didn’t have to worry*) *que somebody te iba a tirar con cerveza o una botella (or something like that)*

“You didn’t have to worry that somebody was going to throw beer or a bottle at you or something like that.”

Spanish/English (Poplack, 1981: 170)

Within Myers-Scotton’s MLF model, *somebody* in this example was identified as an EL content morpheme in an ML + EL constituent, partly because there are no single morpheme constituents in the MLF model (cf. p. 40, 41). As the Monolingual Structure Approach does not exclude such constituents, let us reconsider the possibility of NP insertion.

To classify *somebody* as an embedded function morpheme runs counter to the observed tendency that only content morphemes are embedded in ML constituents. If we decide that this pronoun is a content morpheme we risk expanding the concept of content morpheme. As for the classification of *somebody* as an EL constituent, it should be taken into account that when the ML is the Community Language and the EL is a Superimposed Language, it is in aggregate more common to find embedded content morphemes than EL constituents. In the reverse case however, when the ML is the Superimposed Language, EL constituents from the Community Language are relatively common, while the insertion of Community Language content words is relatively infrequent. (This phenomenon, and the qualifications Community Language and Superimposed Language are discussed in Chapter 11.) In any case pronouns form a special type of NP and, at the most, a special type of content morpheme. Further arguments are needed in order to arrive at a definitive classification. In this case it is relevant that English Subject NPs, and English NPs generally, are an established category of EL constituents in Poplack’s Spanish/English data.¹¹ This is an argument in favour of ranging *somebody* in the above example with NPs, thus avoiding the inflation of the term ‘content morpheme insertion’ for this particular data corpus.

A categorical classification of all pronouns in codeswitching as either EL content morphemes or EL constituents seems precarious: pronouns tend to pattern with the former category in some language pairs, and fall into the latter category in others. Moroccan Arabic/French data yield examples where EL pronouns match with EL constituents. Insertion of Arabic content morphemes in French constituents is rare but Arabic NPs and PPs in French matrix clauses are fairly common as are some types of Arabic pronouns, of which some examples are listed below. (For examples of embedded MA PPs and non-pronominal NPs, see example (41) in Chapter 1, and

¹¹ Another example of an English Subject NP in a Spanish matrix clause is cited in Poplack (1981: 176). Furthermore, her quantification table (1981: 178) shows that 6% of 400 switches are “between noun phrase and verb phrase”, but note that these 6% can, in principle, include Spanish NPs and non-Subject NPs. See also Table 2 in Poplack (1980: 602), where Subject NPs are shown to make up 3.8% of the switches in a larger corpus collected from the same Puerto Rican speech community.

(34)-(38) above.) The following two examples show EL pronominal constituents in a MA/French table conversation on bone surgery recorded by Slaoui (1986). In (48), the first demonstrative pronoun *hada* “this one” is inserted as the direct Object of the verb *relie*. In this sentence, the prepositional complement, the PP *l-hada* “to this one” is an unequivocal EL constituent. In the next example, the indefinite pronoun *ši haža* “something” is embedded. It is conjoined with *une fracture à plâtrer*, *wella à réduire* to form the predicate of the copula construction *quand c’est (...)*.¹²

- (48) *(on met une plaque ici et deux .. et une grosse vis ici ..) qui relie hada*
 we put a plate over here and two .. one big screw over here REL bind’s DEM•M
 l hada
 to DEM•M
 “We put a plate over here and two .. a big screw over here that binds this to this.” MA/French (Slaoui, 1986, Annexe 3:7)

- (49) *quand c’est une fracture à plâtrer, ya-k? wella à réduire wella ši*
 when it-is a fracture to plaster QTAG or to set or INDEF
 haža, ils peuvent le faire ya-k?
 thing they can it do QTAG
 “If it’s a fracture to be plastered, or to be set, you know, or something, they can do it, can’t they?” MA/French (Slaoui, 1986, Annexe 3:4)

On the other hand, we have for instance Ozóg’s (1987) Malay/English data. First and second person English personal pronouns are embedded in Malay structures in a way that is similar to the insertion of (other) English content morphemes. As far as can be judged from the examples in Ozóg’s article, the insertion of full English constituents is largely restricted to time adjuncts (*just now, that day, before Christmas*, etc. [1987: 85]), and NPs of the type [possessive pronoun N], e.g. *my sister* (1987: 84, 85). The use of English pronouns in Malay/English is exemplified below.

- (50) *dia kasi I*
 she give I
 “she gave me” Malay/English (Ozóg, 1987: 84)

- (51) *list tu I*

¹² Although this is formally the case, one could argue that *wella ši haža* “or something” in (49) is actually a discourse marker that indicates the imprecision of the preceding clause.

list DET/DEM I

“my list” Malay/English (Ozóg, 1987: 78)

To summarise this discussion: in some cases single embedded morphemes coincide with higher order constituents with regard to their distributional properties, an example being single adverbs and AdvPs in certain languages, or pronouns and NPs. Here it must be decided whether it is the higher order constituent which is embedded (in, for instance, the finite clause), or the morpheme (in an ML constituent that consists of this EL morpheme alone). The commonly preferred analysis attempts to keep in line with already attested patterns of insertion in the data corpus concerned, as well as in related corpora (e.g. of the same language pair) and in CS generally. As a result, sometimes embedded single morpheme constituents bear closer resemblance to content morpheme insertion, while at other times they are better classified as EL constituents.

2.2.3 Unwarranted generalisations

In 2.2 it was noted that single EL adverbs are insufficient evidence for the insertion of AdvP constituents in Turkish/Dutch, even though they constitute a full adverbial phrase. In addition, the observation that English pronouns and [possessive pronoun N] sequences occur inside Malay finite clauses did not automatically lead to the assumption that NP embedding in general is a codeswitching mechanism in Malay/English. One reason for this conservatism is that content morpheme insertion is a possible alternative analysis; another reason is that single adverbs constitute only one of several subtypes of AdvPs in Dutch, and personal pronouns and [possessive pronoun N] sequences are only two out of many types of NP in English. Thus the problem arises: how many subtypes of a class Q need to be attested as EL elements before one can speak of ‘class Q insertion’. Again, it is difficult to come up with a definitive answer. Without demanding that all categories of class Q are attested as EL forms, we might at least expect instances of the more common categories in monolingual discourse. In any case, generalisations should not obscure relevant differences between codeswitching varieties. With respect to EL constituents, for example, the insertion of constituents that consist of only content morphemes is generally less constrained than the insertion of constituents that include EL function morphemes. Hence CS varieties that include the insertion of complex EL constituents are essentially different from those where constituent insertion is limited to simple constituent types.

Codeswitching with Arabic and French in North Africa illustrates this argument. Embedded French nouns are usually preceded by one of the French definite articles *{l', la, les}* or the indefinite plural article *des*. These articles mark definiteness, gender, and number whenever this is appropriate according to Arabic grammar, cf. generalisation (10) above. Consequently, in all varieties of Moroccan, Algerian, or Tunisian Arabic/French, we find inserted French NPs that include one of these French

articles in addition to one or more French content morphemes. Various examples of Moroccan Arabic/French were cited earlier in this chapter: (1)-(4), (7)-(9), (37) and (38).

The insertion of other types of French NPs, however, is more constrained. The French masculine singular definite article *le*, for instance, is usually replaced by (an allomorph of) the Arabic prefix *l-* (Wernitz, 1993: 188, on Moroccan; Boumans & Caubet, forthcoming, on Algerian). Embedded NPs containing this article occur only in certain varieties, such as the one described by Slaoui (1986), see examples number (5) and (6) above. Also, embedded NPs that include a French possessive pronoun, demonstrative, or the indefinite singular article *un/une* are restricted to particular social and idiosyncratic codeswitching varieties in which the French elements are more varied. (52) below is reproduced from Slaoui (1986).

- (52) $\text{\textcircled{S}emmer-k\ddot{u}m \text{ \textcircled{š}ef-tu \text{ \textit{une} \text{ \textit{vip\`ere} qui est chaude?}}$
 ever-2PL see-2PL INDEF•F viper REL is warm•F
 “Did you ever see a viper that’s warm?”
 MA/French (Slaoui, 1986, Annexe 1:7)

In other varieties of MA/French, the indefinite French Object NP would be rendered as *vip\`ere*, with \emptyset article, *waḥed l-vip\`ere* or perhaps *\textit{\textcircled{š}i vip\`ere}*. Here, *une vip\`ere qui est chaude* occupies an Object NP in the MA clause, whereas in the aforementioned cases there is no systematic NP to NP correspondence. Recall that, as was pointed out on p. 64 above, the distribution of embedded French [definite article-N] parallels the distribution of Arabic [definite article-N] rather than the distribution of Moroccan Arabic NPs in MA clauses. Even more restricted than French nominal constituents that include an indefinite article seem to be those that contain a possessive pronoun determiner. Such French EL constituents were found to be absent in a substantial corpus of Algerian Arabic/French in which EL NPs with French indefinite articles are commonplace (Boumans & Caubet, forthcoming), but they do occur in a corpus of bilingual Moroccan Arabic/French child language reported on by Bentahila & Davies (1994):

- (53) $w \text{ \textit{toutou} ta-ye-dx\ddot{u}l f \text{ \textit{sa} \text{ \textit{maison}}}$
 and Toutou ASP-3-enter in his•F house
 “And Toutou [a doll] goes into his house.”
 MA/French (Bentahila & Davies, 1994: 126)

Thus for North African Arabic/French CS to equate embedded French [definite article N] with NP insertion (as in Nait M’barek & Sankoff 1988) constitutes an unwarranted generalisation. Moreover, the bare statement that French NP constituents are inserted in Arabic clauses obscures structural differences between CS varieties.

2.3 Predictions of the Monolingual Structure Approach

Before proceeding to a discussion of counter-examples, let us first review the predictions that follow from the insertion approach. On the basis of generalisations matrix structures and categories of embedded material are identified. Thus far we have considered finite clauses and lower order constituents, particularly NPs and PPs, as types of matrix structures. Types of EL material considered were lower order constituents in the case of a finite clause matrix, and content morphemes and (in specific cases) inflected content morphemes in the case of a lower order constituent functioning as the matrix. The predictions of the MSA can be summarised as follows:

ML grammatical structure

The Monolingual Structure Approach assumes that each matrix structure originates in the grammar of only one language. For this reason, the internal make-up of the matrix structure must be entirely attributable to the ML grammar. This concerns the presence or absence of constituents and morphemes in the matrix structure and their linear surface order.

EL insertions

Material from another language is said to be embedded in the ML structure only if this material belongs to an established category of EL insertions. An established category of EL morphemes or constituents systematically corresponds to a category of ML morphemes or constituents. In other words, a category-to-category correspondence is assumed, as opposed to a correspondence of individual morphemes or constituents. Systematic correspondence, or CONGRUENCE, relations are inferred from the distributional properties of the EL category in ML structures. Because of this congruence relation, the recognition of embedded elements is not ad hoc but tied in with an observable pattern. Since an EL element represents an ML category in the ML structure, it is as much an expression of ML grammar as the ML elements that participate in the matrix structure.

Distribution and selection of insertions

The fact that the ML determines the distribution of EL insertions does not only affect word order. It also concerns the selectional properties of the matrix structure. If the ML selectional properties differ from those of a similar EL structure the distribution of EL insertions bears evidence of the ML status of the ML.

This argument has been developed in the discussion of the Moroccan Arabic/French examples at the beginning of this chapter, where the distribution of French [definite article N] sequences was shown to be governed by Moroccan Arabic grammar. As a consequence it was concluded that these French nouns and definite articles are embedded in Moroccan Arabic NP constituents. The same applies to the ML on the finite clause level: the selectional properties derive from the ML. The matrix language manifests itself not only in providing constituent order and inflection but also in the specifications that hold for each slot in the syntactic frame (cf. the

discussion of Schmid 1986 in Chapter 1 p. 45). Selectional properties of the ML at the finite clause level have not yet been considered because they are a theoretically valid but not effective criterion for the identification of the ML at that level.

2.4 Counter-examples and limitations

The final stage in the development of the Monolingual Structure Approach consists in investigating the scope of prediction made possible by the concept of matrix language. The present section addresses some data that challenge the Monolingual Structure Approach as outlined in this chapter. Rather than individual examples, the major types of counter-examples will be presented here. Most of the individual counter-examples found in the published codeswitching data or in the Nijmegen corpus of Moroccan Arabic/Dutch can be ranged under one of these classes. The counter-examples to the concept of ML on the constituent level are considered first. Then, in 4.2, we turn to data that challenge the ML on the clause section level. 4.3 is an evaluation of the counter-examples on both levels: do they give us reason to abandon the idea of a matrix language altogether?

2.4.1 Challenges to the ML model on the constituent level

We can conceive of three types of data that challenge the concept of ML in mixed constituents: the occurrence of EL (function) morphemes that cannot be accounted for by ML grammar; the absence of (function) morphemes which are obligatory according to the ML grammar; and a morpheme order that defies ML rules of constituent structure. The first type of counter-example appears in the form of double marking of the same grammatical feature in both the ML and the EL. Missing function morphemes are a relatively common type of counter-example. Manifestations of EL word order are known to occur in embedded content word collocations and in attributive adjectives and numerals.

2.4.1.1 Double function morphemes

Superfluous morphemes in the ML constituent are counter-examples to the claim that the ML on the constituent level governs the distribution of all morphemes. I have already demonstrated that ML grammatical features can be expressed by either ML or EL function morphemes. A common example of an ML feature that is marked by EL morphology is the plural marking of embedded nouns; see Spanish/English and Moroccan Arabic/Dutch examples (15) and (16) above. In Turkish/Dutch, embedded Dutch nouns commonly receive the Turkish plural suffix, as in the case of *jood-lar* “Jews” in example (54) below. In a minority of the cases, Dutch plural nouns are embedded and plurality is marked by a Dutch suffix (55). In some cases, however, we find both ML and EL marking, resulting in one superfluous morpheme; this is shown in the form *polen-lar* in (56).

- (54) *işte jood-lar-ın reger-en yap-tık-lar-1 land-lar*
 so jew-PL-GEN govern-INF do-VERBALNOUN-PL-POSS country-PL
 “So, the countries governed by Jews.” Turkish/Dutch (Backus, 1992: 96)
- (55) *bun-lar-ın verjaardag-en-a git-me-si güzel ol-uyor*
 DEM-PL-GEN birthday-PL-DAT go-INF-3POSS nice be-PROG
 “It is nice that they go to birthday parties.”
 Turkish/Dutch (Backus, 1992: 111)
- (56) *pol-en-lar-a hollandaca ders ver-di*
 Pole-PL-PL-DATIVE Dutch lesson give-PRET
 “He taught Dutch to Poles.” Turkish/Dutch (Backus, 1992: 90)

Needless to say, double marking can only occur in CS varieties where both ML and EL marking are a possible way of marking an ML grammatical feature. Double plural marking is rare in Turkish/Dutch but more common in other data sets. The likelihood of double marking appears to increase when each language marks the same feature in a different manner, for instance, by means of prefixes and suffixes. Moluccan Malay/Dutch shows further examples of double plurals. In the variety of Moluccan Malay nowadays spoken in the Netherlands plural nouns are marked by reduplication of the singular form, which may be a Dutch EL noun.¹³ Sometimes, however, Dutch plural nouns are embedded and reduplicated:

- (57) *kalau minum terus ada di punya pitje-s-pitje-s-nya*
 if drink continue EXIST DET seed-PL-seed-PL-DET
 “If I continue drinking there’ll be seeds.”
 Moluccan Malay/Dutch (Voigt, 1994: 52)

A completely different kind of double marking, involving two numerals will be discussed presently, see example (70).

2.4.1.2 Bare forms

Sometimes ML function morphemes are lacking in mixed ML constituents. This does not constitute a problem for the MSA if there is an EL function morpheme that takes their place to express ML grammatical features, like in the cases discussed in section 3.1.1. However, quite often a required function morpheme is omitted entirely. Following Myers-Scotton (1993b), EL content morphemes that lack obligatory (ML) inflection will be called BARE FORMS. Since function morphemes are expected to

¹³ Reduplication represents the Indo-European type plurality in this variety of Malay. The semantic aspect of diversity, as for example in Standard Indonesian, no longer applies (Tahitu, 1989: 58; Voigt, 1994: 55).

occur in accordance with ML grammar, bare forms constitute counter-examples to the Monolingual Structure Approach.

Dutch nouns in Moroccan Arabic NPs form an example. More often than not, Dutch EL nouns lack the MA definite article prefix in contexts where this prefix is obligatory. Nortier (1990) was the first to describe this characteristic of MA/Dutch CS. Recall that MA requires the definite article after both attributive demonstratives and the indefiniteness marker *waḥed*, as was pointed out above with regard to MA/French CS. The next two examples show ‘bare’ Dutch nouns in these contexts (see also (43) and (44) above). This matter will be examined in detail in Chapter 5.

- (58) ka-y-dir waḥed *-opleiding* dyał ši ḥaža
 ASP-3-do INDEF training of INDEF thing
 “He does some training on something.” MA/Dutch (Hocine)
- (59) u ṣad ka-t-xemmem ṣla eh bezzaf ṣla haduk eh *-moment-en*
 and still ASP-2-think about er much about DEM·PL er moment-PL
 “And you still think about er .. a lot about those moments (..).”
 MA/Dutch (Samir)

In the following Arabic¹⁴/English example, the verb *look* lacks both Arabic inflection and the English third person suffix *-s*. Consequently, the ML on the finite clause level cannot be established for this example.

- (60) bi-y-ṭuul hay el-*tomato look good* - ṭaṣṭii-ni bite
 ASP-3-say DEM DEF-tomato look good give·IMP-1SG bite
 “He says this tomato looks good - give me a bite.”
 Arabic/English (Myers-Scotton, Jake & Okasha, 1996: 33)

In order to establish to what extent EL bare forms challenge the MSA, one should examine the monolingual speech data of the ML as well. After all, it is possible that the omission of certain supposedly obligatory function morphemes is also a feature of the monolingual language or an idiosyncratic speech variety (Sankoff, Polack & Vanniarajan, 1991: 188).

2.4.1.3 Collocations of content words

Along with single embedded content morphemes, we encounter embedded content word collocations. Many of these can be identified as idiomatic expressions. Apart from their co-occurrence, the components of a collocation may show internal structure

¹⁴ The example stems from an unpublished text corpus collected by Okasha. In Myers-Scotton, Jake & Okasha (1996) it is reported that the corpus includes conversations by immigrants to the U.S. of Lebanese, Palestinian and Egyptian origin.

by their ordering relative to each other. This becomes apparent when the word order of the embedded collocation differs from the ML word order. Common types are Noun-Adjective and Verb-Object collocations.

The English words in (61) cannot be formally distinguished from inserted content morphemes, since their placement is in accordance with ML grammar as witnessed by the relative order of noun and adjective as well as the affix *-e-* which comes between attributive adjectives and nouns in Persian (cf. Mahootian, 1993: 70). Even though *expensive watch* is not an idiomatic expression, it may be considered a collocation, that is, the co-occurrence of these two words in English is probably not coincidental.

- (61) bebin če *watch-e- expensive-i* ru moč-eš-e
 look what watch-EZAFE expensive-INDEF on wrist-3SG-is
 “Look what an expensive watch is on his wrist.”
 Persian/English (Mahootian, 1993: 162)

Example (61) contrasts with (62). In the latter case, the relative order of the embedded noun and adjectives must be attributed to English, the embedded language. The absence of an article in the NP *small hand-graphing calculator* can be explained by assuming that the English words are embedded in a Persian matrix NP.¹⁵

- (62) miš-e *small hand-graphing calculator*
 become-3SG
 “It becomes a small hand-graphing calculator.”
 Persian/English (Mahootian, 1993: 153)

Likewise in (63) below, the original Adj-N order of Dutch is preserved; see Eliasson (1995: 49) for Maori/English examples. In (64) the Korean order Object-Verb is observed in *ippal ttakē* “tooth clean” even though the verbal inflection identifies English as the ML on the finite clause level.

- (63) ka-y-dir-u dik *technisch-e school*
 ASP-3-do-PL DEM technical-AGR school
 “They go to that Polytechnic School ..” MA/Dutch (Warda)
- (64) ēnni 's *not ippal ttakē-ing*
 sister is not tooth clean-PROGR
 “Sister’s not cleaning her teeth.” Korean/English (Choi, 1991: 886)

¹⁵ The basic word order in Persian is verb-final. The opposite order in (62) can be explained within Persian grammar as an instance of right-dislocation, according to Mahootian (1993: 154).

- 4) At variance with the MSA an adjective from a Adj-N language occurs pre-nominally while the ML of the constituent has the order N-Adj:

- (68) *ma ci stanno dei smart italiani*
 but there are INDEF•PL smart Italians
 “But there are smart Italians.” Italian/English
 (Di Sciullo et al., 1986: 15)¹⁶
- (69) *ça serait traγiki décision*
 that would-be tragical decision
 “That would be a tragic decision.” Greek/French (Androulakis, 1994: 363)

The data for individual language pairs or data corpora may be more consistent than the overall picture of adjective insertion. However, I must conclude with Santorini & Mahootian (1995, 1996) that from a crosslinguistic perspective the placement of the attributive adjective cannot be attributed to a single factor such as the language of the adjective or the language of the noun. (Note that when we talk about singly embedded adjectives, the language of the noun coincides with the ML and the language of the adjective with the EL.) In other words, the language that sets the frame for the NP constituent is not always capable of imposing its word order on inserted adjectives.

Unlike attributive adjectives, numerals are seldom mentioned as counter-examples to matrix language approaches to CS. Possibly this is caused by the circumstance that most language pairs studied share the word order Numeral-Noun. There are some indications, though, that numerals are similar to attributive adjectives in having a tendency to retain source language word order when singly embedded. Van Staden (forthcoming), for instance, discusses North Moluccan Malay numerals in Tidore, a non-Austronesian language of the North Halmahera language family spoken on the islands of Tidore, Maitara and Mare and in some villages on Halmahera. In Malay numerals precede the noun whereas they follow the noun (and adjective) in Tidore. When Malay numerals are used in Tidore sentences they sometimes (but not always) precede the counted noun. In the following example the ML is Tidore according to Van Staden’s criteria¹⁷ and the numeral “seven” surfaces twice: prenominally in Malay and post-nominally in Tidore. Note that this is another instance of double marking.

¹⁶ The pre-nominal position of the adjective is fine in Italian if it is not used in a restrictive sense. Whether this applies to the present example cannot be established as it is presented out of context.

¹⁷ Both Tidore and Malay have little inflectional morphology. In the Tidore/Malay case, the inflection of the finite verb can therefore not be applied as a criterion to identify the ML in the finite clause level. According to Van Staden, “native speaker judgments show that if functional elements, in particular bound morphemes, in a sentence are Tidore, then the entire sentence will be considered Tidore. In particular, demonstrative pronouns, negators and affixes are good indications of the matrix language.”

- (70) *turus so-masusu toko toma kurunga, turus kage yali tulju wange*
 so CAUS-enter chicken LOC cage so moment more seven day
tomdii
 seven
 “So then [he] put the chicken in the cage, and then it was seven days later.”
 Tidore/North Moluccan Malay (Van Staden, forthcoming)

Van Staden notes that similar examples are not common in her data.

2.4.2 Challenges to the ML model on the finite clause level

The ML on clause level has been defined as the language of the inflection of the tensed verb, and the main prediction was that the relative order of the verb and its arguments can be attributed to this ML. Therefore counter-examples consist in the first place of cases where verbal inflection and constituent order cannot be ascribed to the same language. Subsequently, I will address the relevance of the ML at the finite clause level for the placement of EL non-argument constituents.

2.4.2.1 Inflection and constituent order cannot be ascribed to the same language

Stenson (1990) cites one example, reproduced here as (71), of an English inflected verb followed by an Irish Subject NP. The Verb-Subject order argues for Irish as the ML, while the verbal inflection is English. Stenson (1990:173) reports that there is only one such example in her Irish/English data set, all other English (EL) verbs being either morphologically assimilated (see (31)) or infinite forms (participles).

- (71) *decided aer lingus go raibh sé ro-chancy*
 decided Aer Lingus that be-PA it too-chancy
 “Aer Lingus decided that it was too chancy.”
 Irish/English (Stenson, 1990: 174)

Nishimura (1986) cites a number of Japanese/English clauses in which the inflected verb is English while the constituent order seems to be Japanese:

- (72) *mannaka ni they’re growing*
 middle in
 “They’re growing in the middle.” Japanese/English
 (Nishimura 1986: 132)

- (73) *asoko she goes*
 that:place
 “She goes over there.” Japanese/English (Nishimura, 1986: 132)

In (73) *asoko* has a directional meaning according to Nishimura, although the Japanese directional marker *ni* or *e* is deleted (1986: 132). Because of this, (73) cannot be interpreted as “There she goes”. Furthermore, pre-posed PPs are possible in English as topicalised (‘thematic’) constituents, but, the author argues, the context of (72), (73), and similar examples excludes such a reading (Nishimura, 1986: 133). Instead, she proposes a Japanese sentence structure for these examples, with the English VP as an embedded Predicate constituent, see Fig. 2.2. The Japanese constituent is the Topic in this sentence. Japanese has a Topic marker *wa*, but this is omitted in some contexts in spoken discourse (Nishimura, 1995a: 40), as in (72) and (73). Such an analysis is problematic for the Monolingual Structure Approach because, by definition, an EL constituent cannot include the finite verb inflection, as do *they’re growing* and *she goes*. However, the author revises her position somewhat in a subsequent article (Nishimura, 1989). She points out that the relationship between Topic and Comment in (monolingual) Japanese is not a syntactic one, but rather a matter of discourse organisation. This means that the initial Japanese constituents *mannaka ni* and *asoko* are outside the following English finite clause, and the MSA is saved.

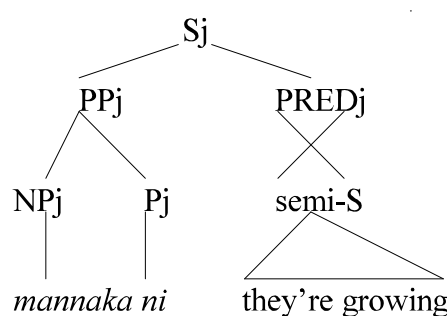


Fig. 2.2. Japanese Topic-Comment structure (Nishimura, 1986: 133)

2.4.2.2 The placement of non-argument constituents

Thus far I have only discussed constituent order with respect to the verb and its arguments. The next question is: how does the ML relate to the placement of non-argument constituents, in particular adverbials and adjunct PPs? The placement of adverbs and other non-argument constituents has received little attention with regard to CS, with the exception of discourse markers that are considered to be outside the syntactic structure (‘extra-sentential CS’). Adjunct PPs, such as locative and temporal adjuncts, seem to comply with the predictions of the MSA, as do place and time adverbs. The case of adverbs that express aspect or subjective modality is different, however.

Placement rules for this type of adverbial constituent tend to be both variable and complicated, and, therefore, more poorly described even as far as the monolingual language varieties are concerned. This creates difficulties in identifying counter-examples to the claim that their syntactic position is governed by the ML. However, Stenson (1990: 173, 182-3) notes that a number of English adverbs (viz. *still*, *already*

and *definitely*) occur in ‘medial’ position when embedded in Irish clauses, cf. (74), whereas their Irish counterparts are far more common in clause-final position.¹⁸

- (74) beidh mé thart *definitely* anocht
 be-FUT I around definitely tonight
 “I’ll be around definitely tonight.” Irish/English (Stenson, 1990: 173)

Stenson’s observation concerning particular EL adverbs in Irish/English is confirmed by some similar observations on modal and aspectual adverbs in other language pairs, which will be discussed in the next chapter on discourse organisation. However, it is not clear to what extent these facts can be generalised. At this point I can only say that there is no unequivocal evidence that the ML as the language of the verbal inflection governs the placement of non-argument constituents, while there is some evidence to the contrary. Since the ML does not consistently make the right predictions about the placement of (types of) adverbial constituents, this must be recognised as a limitation of the ML model.

It is worthwhile investigating the characteristics of adverbs that display source language behaviour. The distribution of EL adverbs in ML clauses may be associated with certain pragmatic functions, with properties of the ML or the EL or with sociolinguistic variables of the language contact situation at stake. As already mentioned, since the word order properties of adverbs are often rather complicated it will not be an easy task to determine the mechanisms governing EL adverb placement. Note, however, that such questions will prove even more difficult to study if these counter-examples induce us to reject the concept of ML altogether.

2.4.3 Adjustment of the Monolingual Structure Approach

Now that various counter-examples have been presented which have all been argued to represent categories of counter-examples I will evaluate how the model can be manipulated so as to incorporate these challenging data.

To abandon the idea of a matrix language altogether would be throwing out the baby with the bathwater. After all, the evidence that was presented in favour of a matrix language approach is still valid. Moreover, we would be ignoring the telling fact that the challenging data come in recognisable patterns rather than in an array of all kinds of counter-example. Instead of abandoning the MSA, we will investigate possible explanations for the CS patterns that defy the model.

¹⁸ With regard to the discussion on single morpheme constituents in section 1.3.2 above, note that the analysis of *definitely* and other EL adverbs as EL constituents is preferred here. To assume that the EL adverbs are embedded in an ML adverbial phrase would engender the extra complication of an ML constituent that defies ML constituent order.

Many cases of double marking, for instance, are isolated instances. Their explanation may be the same as for occasional instances of double marking in monolingual speech. Double marking is more likely to occur with two different markers, preferably also of a different kind (e.g. a prefix and an affix, or a prenominal and a post-nominal numeral like in (70)). This probably makes double marking more common in CS than in monolingual speech where double marking would often involve the repetition of the same marker. (For an example of monolingual double marking see the ungrammatical repetition of the negator *niet* in the Dutch clause in example (14) on p. 308.)

Some of the bare forms discussed in 4.1.2 may be explained by a relative or absolute unproductivity of the ML morphological rules that are needed to provide the missing ML function morphemes. The fact that it is mainly affixal morphology that is missing supports such an explanation. Affixes are not equally productive in the sense that they can combine with new or, in the case of CS, foreign stems to form new word forms: some affixes are not productive at all, while others are sometimes, but not universally, applied to new words. Language-internal, independent arguments are possible in support of this claim. The MA definite article in MA/Dutch is an example of ‘semi-productive’ morphology: the article *l-* is prefixed to Dutch nouns in some cases, but mostly it is not. This is similar to the behaviour of this prefix in monolingual MA, where it is often omitted before loanwords from Berber or Romance languages. See Chapter 5 for more details.

Concerning the remaining data challenging the MSA there are two adjustment strategies. One is to relax the model by admitting other types of insertions, the other is to limit the applicability of the Monolingual Structure Approach in its present configuration.

2.4.3.1 Other categories of embedded material

In section 1.2.1 above we have seen that, besides content morphemes, inflected and derived word forms can be embedded in ML constituents, and that EL function morphemes can function as an expression of ML grammar. At this point we may also contemplate the possibility that the finite verb, which serves to identify the ML on the finite clause level, is itself an insertion, in order to explain counter-examples like the Irish/English example in (71). Although highly uncommon, occasionally there is convincing evidence of such an insertion. This applies to those cases where it can be established that the EL inflected verb, given its distribution and paradigmatic organisation, expresses ML verbal categories. Igla (1991), for example, reports on a dialect of Romani nowadays spoken in a suburb of Athens in Greece by a community that moved in from Turkey in the 1920s. The speakers of this dialect of Romani no longer speak any Turkish, however there are still approximately 30 verbs of Turkish origin which continue to be inflected with Turkish suffixes while following the Romani verbal paradigm. That is, the Turkish inflectional suffixes express Romani inflectional categories (cf. Boumans, forthcoming). A famous case in point is Mednyj Aleut, also known as Copper Island Aleut (cf. Thomason & Kaufman, 1988: 233-8).

In this variety of Aleut, Aleut inflectional patterns in finite verbs were replaced by Russian ones while most other grammatical subsystems remained intact. Crucially, the Russian verbal inflections express Aleut tense and aspect categories. Just like the variety of Romani described by Igla, Mednyj Aleut is a stabilised speech variety as opposed to a form of CS, although its present form may be historically related to a CS variety which displayed the same adventitious combination of Aleut verb stems and Russian inflectional affixes. Naturally, in the case of Mednyj Aleut/Russian CS the finite verb inflection could not be employed as a reliable criterion for the identification of the ML. The possibility of the finite verb being an insertion could explain example (71) above, although, in this case we lack evidence that *decided* expresses Irish verbal categories.

Obviously, the admission of the possibility that the finite verb itself is an insertion drastically damages the MSA as outlined in this chapter, since it is precisely the finite verb which designates the ML. Because of this, the argument that the finite verb is itself an EL form cannot be invoked. In the case of compelling evidence that inflected EL finite verbs express ML verbal categories, the phonological shape of the finite verb, i.e. the ‘lexeme’ in the terminology used by Levelt (1989) for his speech production model, is no longer a valid criterion for the identification of the ML. Instead the semantic and pragmatic properties of the verb inflection would be decisive in determining the ML, however this is admittedly an unworkable criterion in CS contexts. Similarly the finite verb criterion cannot be used to identify the ML on the finite clause level in the case of CS between two languages which both lack verbal inflection. In such cases other criteria must be invoked to identify the ML, for instance constituent order or other function morphemes which appear to consistently coincide with the syntactic structure of the clause, see Van Staden (forthc.). However, with most CS varieties studied in the literature the inflectional element of the finite verb appears to be a reliable and practical criterion.

Collocations constitute another category of EL entities that could be admitted in order to save the MSA. Consequently, the preservation of EL-specific word order in embedded Noun-Adjective and Verb-Object collocations would no longer be viewed as a counter-example to the model. The preservation of their internal structure could be explained by assuming that collocations, like compound words and some derived and inflected forms, are stored in the speaker’s lexicon in a fixed format.

2.4.3.2 Limitations of the MSA: Recalcitrant word order

To acknowledge the limitations of the Monolingual Structure Approach as developed so far is the only possible reaction to the facts that challenge the MSA with respect to word order. The conclusion that the ML as it is defined does not govern all aspects of word order comes as no surprise when we realise that the word order in the clause “results from the interaction of a number of independent principles”, as Muysken rightly observes. Muysken (1995: 195) mentions a number of such principles, which are reproduced here:

- (75) directionality of government (Case, Theta)
[NP V], *[V NP] under leftward government
[P NP], *[NP N] under rightward government
- (76) adjacency or other locality conditions on government
[V NP X], *[V X NP], since case assignment is local
- (77) iconicity
[E1 E2], *[E2 E1], where E1 and E2 are coordinate events and E1 preceded E2 in time
- (78) considerations of given/new, functional sentence perspective, topic/comment, etc.
[given information new information]
- (79) prosodic considerations
[short constituent long constituent]

It is possible that some of these principles, and other ones not mentioned here, are regularly associated with the ML as defined in (14) and (30) while others are not.

The ‘directionality of government’ principle in (75) could account for the tendency of Verb-Object collocations to preserve EL word order, as described above in 4.1.3, although Verb-Object collocations do not invariably show EL order (see, for instance, example (23) in Chapter 1). Regarding verbs, this word order principle interferes with the definition of the ML at the finite clause level. Concerning adpositions, the ‘directionality of government’ principle is incorporated in the characterisation of ML adpositional phrases. Thus the PP *basement de* in (33), for instance, was analysed as a Japanese matrix constituent partly because of the word order.

The working of principle (78) on mixed Japanese/English utterances has been demonstrated by Nishimura (1986, 1989), who asserts that the Japanese Topic-Comment structure is not a syntactic structure (subsection 4.2.1 above).

The whimsical behaviour of EL attributive adjectives in mixed NPs must result from an interaction of ML and EL word order principles. It is possible that the word order is related to the adjective’s descriptive, specifying, or contrastive function in an individual instance. If this could be established, we could conceive of an explanation which associates some of the pragmatic functions of the adjective with the ML and others with the EL. Languages such as Spanish and Italian, where the relative order of noun and attributive adjective depends on the adjective’s communicative function, point to the possible relevance of such considerations. Stenson’s (1990) investigation of EL attributive adjectives in Irish/English demonstrates the relevance of pragmatic factors. Stenson discovered that embedded English adjectives follow Irish (ML) post-nominal word order, except for what she calls expletives, i.e., words like *fuckin’*, *friggin’* and *bloody*. These occur in pre-

nominal position as in English. An example of a ‘normal’ English adjective has already been cited above in (66); a further example is presented below and an ‘expletive’ adjective is shown in (81).

- (80) tá carr *light green* aige
 be car light green at-him
 “He has a light green car.” Irish/English (Stenson, 1990: 171)

- (81) cá bhfuil mo *fuckin* sheaicéad?
 where be my fucking jacket
 “Where’s my fuckin’ jacket?” Irish/English (Stenson, 1990: 171)

The Irish/English case exhibits an unequivocal relationship between EL word order and the adjective’s communicative function: unlike the adjectives in (66) and (80), ‘expletives’ do not semantically modify the head noun, but express the speaker’s appreciation of some state of affairs, e.g. being unable to find one’s jacket.¹⁹ The examination of patterns of adjective insertion is complicated by the fact that embedded attributive adjectives occur very rarely in most CS corpora. I will have more to say about adjectives in Chapter 11.

It turns out that only certain aspects of word order relate to the present definitions of ML, that is, to ML constituent assembly and to finite verbal inflection. Aspects of word order in mixed constituents or clauses that are at odds with ML syntax must be attributed to other principles of word order that are probably associated with the other language. The relative order of the Subject and the finite verb, in particular, relates to the finite verb, but the relative order of the verb and its complements may eventually be more directly related to the ‘directionality of government’ properties of the lexical verb. However, I do not propose to identify an ML for each word order principle. Instead, I propose to stick to the ML on the finite clause level and on the constituent level as defined so far, and to use these matrix structures as an anchor for the identification of EL principles of word order that bear on the mixed structure.

From the perspective of the Monolingual Structure Approach, EL word order phenomena can be dealt with in three ways. One way allows for embedded elements to bring along EL word order properties despite their being embedded. This solution, which entails a relaxation of the model, will be adopted for the embedded Verb-Complement and Adjective-Noun collocations, as well as EL sentence adverbs like *definitely* in the Irish/English example (74). The second option is to relate the non-ML word order phenomena to a supra-clausal matrix structure, such as the Sentence or Complement Phrase in the other language. Finally, the non-ML word order

¹⁹ Stenson (1990: 177) reports one other prenominal embedded English adjective: *extra*. Perhaps its prenominal position can be explained by associating it with quantifiers or determiners rather than adjectives.

phenomena may be located at an extra-clausal level that has no hierarchical relation to the ML clause. A non-syntactic matrix structure like Topic-Comment in Japanese exemplifies this solution. All three options will be reconsidered in Chapter 2 with respect to elements that function on the level of discourse organisation.

2.5 Summary

In this chapter I elaborated on the approach that analyses codeswitching in terms of a matrix structure from one language in which elements from another language are embedded. A matrix frame was recognised on two levels: on the finite clause level and on the level of lower order constituents, simply called the constituent level. On both the constituent and the finite clause levels, the occurrence of EL elements was restricted to established insertion categories. Generalisations over a set of CS data identify categories of EL material as possible insertions.

Constituents as matrix structures

On the constituent level the ML was defined as the language that can account for the occurrence and the relative order of the morphemes that make up the constituent. EL content morphemes are commonly embedded in the ML constituents. In addition, some types of EL function morphemes are regularly embedded together with their content morpheme head, EL plural nouns being the most common example. It was shown that attached EL function morphemes may express ML grammatical functions. Because of the possible insertion of derived and inflected content morphemes, not all function morphemes necessarily derive from the ML.

ML finite clauses

On the clause level, the ML was defined as the language of the inflection of the finite verb. It was claimed that the ML determines the distribution of constituents in the finite clause, where distribution implies both the occurrence and the relative order of the constituents that make up the finite clause. It was shown that the ML thus defined efficiently accounts for the constituent order of mixed finite clauses in a variety of language pairs, in particular for the relative order of the verb and its arguments.

Layered insertion

Since the ML is defined independently on more than one level, it is possible to have successive layers of insertion. For example, an English noun can be part of a Japanese PP, which is in its turn embedded in an English clause. In this way, many cases that are problematic for other matrix language models receive a more satisfactory explanation. The possibility of layered insertion must be recognised as a logical corollary if the insertion of content words and of (complex) constituents is permitted.

Counter-examples

Some patterns of counter-examples were pointed out, in which none of the languages involved could entirely account for the make-up of the constituent or the finite clause. At the constituent level, some embedded content words occasionally or regularly lack affixes or other function morphemes required by ML grammar. In other cases, both ML and EL marking surface, resulting in one superfluous morpheme. When collocations of EL content words are regularly embedded, the relative order of the content words that make up the collocation is often preserved even if this order differs from the ML word order. EL attributive adjectives are often found to follow EL order when inserted into ML noun phrase constituents.

The shortcomings of the MSA were also pointed out at the finite clause level. There is some evidence that modal and aspectual adverbs tend to retain source language word order properties when they occur as EL elements. Similarly at this level, collocations of EL content words, notably Verb-Object collocations, challenge the MSA if they retain EL word order.

The Monolingual Structure Approach as outlined here summarises a number of regularities and recurrent patterns of codeswitching in an orderly way. This is a major advantage of the MSA: it provides a uniform framework for the description of code-switching data from various language pairs and sociolinguistic contexts. Despite the observation that the ML cannot predict all regularities it is still more economical to assume an ML to which the large majority of the patterns can be ascribed and to state the exceptions separately, than to follow a left-to-right linear approach.

On the other hand, the Monolingual Structure Approach must be credited for making the right predictions in the majority of cases. The fact that CS can be described according to this model is not self-evident: not every random juxtaposition of elements from two languages would fit in. Now that a large part of the regularities in CS behaviour can be explained with reference to ML grammar, the manifestations of the other language, called EL, can be easily distinguished and investigated. More predicaments for the MSA will be investigated in the following chapter which focuses on discourse grammar. After that I will apply the MSA to the Nijmegen corpus of Moroccan Arabic/Dutch and demonstrate how data description works in this framework.

Chapter 3

Discourse Grammar

In the preceding chapter the ML at the finite clause level was defined as the language of the tensed verb. This implies that, while Subject and complement constituents may be realised in the EL, their position in the ML clause structure is in accordance with ML grammar. Indeed, the placement of EL core constituents as observed in a number of language pairs was instrumental in the establishment of both the notions of matrix language and constituent insertion. Now what about the placement of elements less central to clause syntax such as adverbial adjuncts, particles, and conjunctions?

Even though the identification of an ML carries the suggestion that all elements are placed and ordered according to ML grammar, and despite other researchers having used the ML concept in this sense, there are indications that such a claim cannot be upheld with regard to all word classes. For various language pairs researchers have shown that the occurrence and word order properties (i.e., the distribution) of certain adverbs, particles, conjunctions and (other) discourse markers reflects the syntax of their source language even when these occur within ML finite clauses from the other language. What these elements seem to have in common is a function on the discourse level, either in terms of sequencing units of talk or as expressions of evaluative modal values. In the course of this chapter I will discuss examples involving adverbs, conjunctions and particles that function as discourse markers as well as some other adverbs which function as markers of tense and aspect.

In discussing possible counter-examples to the MSA in Chapter 2, I already touched on another aspect of discourse grammar: apparent counter-examples from Japanese/English were discarded as instances of an extra-clausal foregrounding construction known as the Topic-Comment construction in Japanese grammar. Although this foregrounding construction, by virtue of being extra-clausal, does not constitute a real threat to the ML on the finite clause level, the MSA does not succeed in describing or explaining this aspect of CS either. In the following we will take a closer look at both extra-clausal foregrounding constructions and discourse marking as manifestations of discourse grammar, and examine their interpretation within the Matrix Language approach to CS.

The functions of discourse grammar can be understood in terms of two broad categories: sequencing and evaluating. A number of very diverse mechanisms mark the sequencing of stretches of discourse. One is the introduction or re-introduction of a referent as topic by means of a foregrounding strategy. In this chapter two extra-clausal foregrounding constructions will be addressed in some detail: left-dislocation and topicalisation. Another aspect of sequencing concerns the ordering of sections or episodes in a text such as the chronological ordering of events in a narrative, and

various causal relations. This ordering of sections in a discourse can be signalled by discourse markers, typically called conjunctions. The marking of turn-taking, floor-keeping or floor-yielding may be considered yet a further aspect of sequencing in conversations. The second broad category of discourse grammar concerns the speaker's expression of her attitude toward what is being said or toward her interlocutors. Such attitudes are often expressed by modal adverbs and clause-final tags.

Before we pass on to the discourse grammar of CS it is important to note that even in monolingual contexts the syntactic distribution of discourse markers cannot be fully explained within sentence grammar since they function entirely or partly on the level of discourse organisation (Stubbs, 1983: 67-83; Schiffrin, 1987: 37-40). Related to this is the problematic classification of various markers into word classes such as conjunctions, adverbs, interjections or simply discourse markers. The distinction between these classes is not always clear-cut and each word class tends to be rather heterogeneous in itself. Moreover, the classification in terms of syntactic properties does not necessarily reflect different pragmatic functions.¹

I will first present Schiffrin's (1987) definition of discourse markers and illustrate the point with an example from the Nijmegen codeswitching corpus (section 1). Section 2 gives an overview of the problems the MSA encounters in dealing with prototypical discourse markers such as English *well*, sentence adverbs, and conjunctions. In this section special attention will also be focussed on clause-external foregrounding strategies, in particular the use of so-called discourse emphatic pronouns, which is so prominent in spoken Arabic and, consequently, in CS with Arabic. Section 3 examines the possibility of accounting for these discourse level phenomena within a syntactic model with emphasis on the proposals by Eid (1996) and Myers-Scotton, Jake & Okasha (1996) to explain Arabic discourse emphatic pronouns in terms of X-bar syntax. At the end of this section some alternative explanations receive consideration as well. In the final section I will recapitulate the arguments presented in this chapter and present some concluding remarks.

3.1 Defining discourse markers (Schiffrin 1987)

Discourse markers form a heterogeneous group of expressions that include members from various word classes. English examples include adverbs like *now*, sentence adverbs like *frankly*, *firstly*; coordinate and subordinate conjunctions like *and*, *because*; 'particles' such as *well*, *right*, interjections like *gosh*, and even complete clauses, e.g. *you know*, *I see*. Other categories (and taxonomies) can easily be added to this list. The linguistic properties are at least partly dependent on their word class

¹ See, for instance, De Vriendt & Van de Craen (1993) for a concise discussion of Dutch conjunctions, adverbs, 'conjunctive adverbs' and 'conjunctive particles'.

membership, and because of this it is difficult to come up with a clear definition of discourse markers.

Schiffrin's study of discourse markers in American English offers the 'operational definition' of discourse markers as "**sequentially dependent** elements which bracket units of talk" (1987: 31). The author then proceeds to work out the core elements of her definition. After discussing the limitations of more precisely defined speech units, she concludes: "Markers bracket units of talk. Sometimes those units are sentences, but sometimes they are propositions, speech acts, tone units" (1987: 35). Related to the intricacy of defining the relevant units of talk in terms of linguistic features is the concept of sequential dependence: "Markers are devices that work on discourse level: they are not dependent on smaller units of talk of which discourse is composed" (1987: 37). The notion of sequential dependence is also intended to convey the anaphoric as well as the cataphoric character of discourse markers, that is "brackets look **simultaneously** forward and backward - (...) the beginning of one unit is the end of another and vice versa" (1987: 37). Schiffrin concludes her study with a more theoretical definition: "markers provide **contextual coordinates** for utterances: they index an utterance to the local contexts in which utterances are produced and in which they are to be interpreted" (1987: 326).

Note that these definitions do not exclude the functioning of discourse markers at sentence level, but are intended to include markers that defy an interpretation within sentence grammar. To illustrate the latter category, an excerpt from the Nijmegen corpus of Moroccan Arabic/Dutch codeswitching data is quoted below.

This passage shows a conversation between Samir, who was doing the recording, and three young women, Zineb, Warda, and Maryam. In the beginning of the recording session, Samir tries to collect sociolinguistic information on language use from his interlocutors.

(1) Sociolinguistic inquiry with Samir, Zineb, Warda, and Maryam

Samir sakn-a sa.. sakn-a b ruḥ-ek wella sa.. eh sakn-a mʃa
live·PART-F liv.. live·PART-F with self-2SG or li.. er live·PART-F with

l-walid² dyal-ek?

DEF-parents of-2SG

"Do you live alone or do you live with your parents?"

Zineb sakn-a mʃa l-ʔax dyal-i
live·PART-F with DEF-brother of-1SG
"I live with my brother."

² One of the idiosyncrasies of Samir's MA is the use of singular *walid* "father" to mean "parents" (*walid-in*).

Samir l-ʔax?

DEF-brother

“Your brother?”

[SAMIR JOTS DOWN SOME NOTES; THE WOMEN START LAUGHING AND GIGGLING]

Zineb *aan 't kladd-en?*

on DEF scribble-INF

“Taking notes?”

[LAUGHING AND GIGGLING CONTINUES AND SAMIR JOINS IN]

Samir *muhiimm, dat zijn eh, dat zijn eh dingetjes eh vragen die ik gewoon*
 anyway those are er those are er whatsit er questions that I just

moet weten, la-xateɽ op een gegeven moment moet ik eh, zij moeten
 must know because at a given moment must I er they must

weten šhal men ʕam ʕend-ki u šhal men ʕam ɾa-ki
 know how-many of year at-2F and how-many of year PRES-2F

sakn-a hnaya u mʕa men sakn-a, sakn-a b waħd-ek wella
 live-PART-F here and with whom live-PART-F live-PART-F with one-2SG or

mʕa l-walid dyaɛ-ek, dat zijn allemaal belangrijk-e
dingen om eh
 with DEF-parents of-2SG those are all important-AGR things for er

te weten, waš ʕend-ki ši l- eh l-ʕalaq-at mʕa l-ʔašdiqa?
 to know Q at-2F INDEF DEF- er DEF-relation-PL with DEF-friend-PL

huɛandi-yin, wella l-ʕalaq-at mʕa l-mɣarba?
 Dutch-PL or DEF-relation-PL with DEF-Moroccan-PL

“Anyway, those are er, those are er whatsit er, questions that I just have to know. **Because** at a given moment I have to er .. THEY have to know how old you are, how many years you have been living here and with whom you are living, whether you are living on your own or with your parents .. Those are all important things to know. Do you have relationships with Dutch friends, or relationships with Moroccan friends?”

When Zineb, Maryam and Warda start laughing, Samir realises that asking a woman whether she lives on her own is an unconventional move in the present interaction context. He then feels obliged to explain that his questions are meant to serve academic purposes. Although MA *la-xateɽ* “because” in the second line of Samir’s third turn is traditionally classified as a clause conjunction, in this example its function is located on the discourse level. Here *la-xateɽ* does not introduce a subordinate clause that contains an argument in support of a proposition stated in the main clause. Rather it functions as a marker of what Schiffrin (1987: 25) calls the

action structure of the present discourse: *la-xaṭer* introduces a relatively elaborate justification of Samir's prior action (viz. posing impertinent questions). In fact, the use of causal conjunctions to signal the justification of a speech act ("I ask/beg/insult you, BECAUSE") rather than factual causal relations is quite common in informal discourse, at least in some languages. We return to this in 3.3.2 below.

Admittedly this introduction can provide no more than a vague impression of the vast domain of discourse marking and discourse grammar. In the next section further examples are discussed in relation to CS.

3.2 Challenges to the matrix language approach

Conjunctions in particular have often been cited as counter-examples to matrix language models which, like Myers-Scotton's MLF model, predict that function (system) morphemes originate from the ML (e.g. Eliasson, 1995: 56; De Rooij, 1996: 161). There are of course valid arguments supporting the classification of conjunctions with function rather than content morphemes: their use is better explained in terms of (discourse) functions than in terms of referential meaning. It is perhaps in reaction to such criticism that, in the recent versions of her model, Myers-Scotton claims that discourse markers are content morphemes because they assign thematic roles on the discourse level (Myers-Scotton & Jake, 1995: 984; Myers-Scotton, Jake & Okasha, 1996: 13).³

In principle, the same reasoning for and against system morpheme status could be applied to other discourse markers like *yes*, *all right* and *please*. But the question simply does not pose itself here, since such markers mostly occur either clause-initially or clause-finally: they are not traditionally studied in sentence grammar, nor are they considered intra-sentential codeswitching (see extra-sentential CS, Chapter 1, p. 11).

In this section I will show that the ML on the finite clause level as it is defined in the preceding chapter, viz. the language of the finite verb, does not make solid predictions on morphemes that function on the discourse level. The ML does not predict from which language discourse markers will be drawn, nor does it predict anything about the distribution of these markers. More generally, discourse grammar interferes with the finite clause level in such a way that it cannot be handled satisfactorily within clause syntax. Besides discourse marking particles, adverbs and conjunctions I will also address clause-external foregrounding constructions.

³ In her 1993 book, however, Myers-Scotton proposes another explanation for the 'troublesome' occurrence of the French complementizer *que* in otherwise Lingala contexts, viz. that "*que* has become an established borrowed form in Lingala and therefore is accessed as a Lingala system morpheme" (1993b: 132).

3.2.1 Discourse marking particles

Discourse markers form a significant portion of CS. Since many discourse marking particles do not display clear syntactic properties, they generally do not challenge the ML concept, but there is hardly any indication that these are embedded elements either. A few cases are presented here.

3.2.1.1 Japanese/English

Japanese has sentence-final particles that also occur in largely English contexts in Japanese/English (Nishimura, 1989, 1995b). Examples are *ne*, to call for confirmation, as in (2), and *yo*, to express conviction, an example of which is found in (30) below.

- (2) *She's the one that borrowed it ne*
 "She's the one that borrowed it, right?"
 Japanese/English (Nishimura, 1995b: 168)

Since sentence-final tags (*right? you know*) are common in monolingual English, these particles might be considered EL forms. They only challenge the ML concept if it can be shown that English has no discourse markers with the same function and distribution, in which case insertion is not a feasible explanation. A more elaborate comparison of Japanese and English discourse markers is required in order to decide whether these Japanese discourse markers can really be considered embedded elements. A factor weighing against the consideration of these particles as EL forms is that, as function morphemes, their embedding is an unlikely possibility.

3.2.1.2 German/English

Salmons (1990) reports on English origin discourse markers that have become an established part of the discourse marking inventory of Texas and Indiana German dialects. *Well* and other markers "appear in a distinctly English distribution and one very much unlike the German modal particles, by virtue of occurring utterance-initially but not utterance-finally" (Salmons, 1990: 467). Since the English markers occur in contexts where the ML, as defined in the MSA, is German, the ML cannot account for their distribution, cf. (3).

- (3) *wos ich gesagt he? oh well, die sin do herkomme, wo ..*
 "What I said? Oh well, they came from there, where .."
 German/English (Salmons, 1990: 459)

3.2.2 Adverbs

Various adverbs have been reported to maintain their source language word order properties when they are part of a finite clause in another language. Most cases

concern sentence adverbs that express the speaker's orientation toward what is being spoken of (*fortunately, already*), or mark sequences in the discourse (e.g. *first(ly), finally, suddenly*). Sentence adverbs modify entire sentences rather than verbs, adjectives or other adverbs. They are distinguished from adverbs expressing time (*yesterday*), degree (*extremely*), place and direction (*home*) or manner (*slowly*); see Schachter (1985: 20) for this classification of adverbs. While most adverbs that challenge the insertional approach to CS express discourse sequencing and subjective modality, others involve the expression of objective modality, tense/time, and aspect.

In many languages sentence adverbs occur clause-initially or clause-finally, so it is difficult to find instances that challenge the Monolingual Structure Approach on syntactic grounds. In the following, I examine cases where the languages involved in CS differ from each other with respect to the syntactic properties of sentence adverbs.

3.2.2.1 English and French adverbs in German and Dutch

Salmons (1990) reports a tendency for German subject-verb inversion, i.e. the 'verb second' rule, not to apply when the first element is an English discourse marker, see (4). In this example, the placement of *of course* is at odds with the ML syntax, as the finite verb appears in third position.

- (4) und *of course* ich konnte kein *french* oder italienisch
 and of course I could no French or Italian
 "And of course I couldn't speak French or Italian."
 Texas German/English (Salmons, 1990: 461)
 cf. Standard German: "Und natürlich konnte ich ..."

Likewise most French adverbs in Treffers-Daller's Brussels Dutch/French data do not trigger the 'verb second' rule in Dutch. Treffers-Daller (1994: 179) counts only three tokens where French adverbs ARE followed by the finite verb in Dutch (see (5) below), as against 103 adverbs in a sentence-initial position that do not trigger subject-verb inversion, as in (6), and 318 tokens following the finite verb. As far as I can verify, the 426 instances considered by Treffers-Daller concern sentence adverbs associated with discourse marking functions, except for 29 unspecified tokens of adverbs ending in *-ment* which could be manner adverbs. This distinction is relevant because manner and sentence adverbs have different syntactic properties, which will be discussed presently. The case of *automatiquement* in (5) is ambiguous. It can be interpreted as a sequencing adverb, marking a new episode in the narrative. In this case, it constitutes an exception to the rule that French sequencing adverbs do not trigger the 'verb second' rule. Alternatively, although it does not indicate the manner of speaking, it could be interpreted as a manner adverb modifying the implied meaning of "to switch to Flemish". The latter interpretation would explain the occurrence of subject-verb inversion despite the adverb being in French.

- (5) *en automatiquement* klapte gij ook schoon vlaams
 and automatically spoke you also standard Flemish
 “And automatically you would switch to standard Flemish.”
 Brussels Dutch/French (Treffers-Daller, 1994: 178)
- (6) *d’ailleurs* ’t gasthuis heeft ’t ook geconfirmeerd
 anyway the hospital has it as-well confirmed
 “Anyway, the hospital has confirmed it as well.”
 Brussels Dutch/French (Treffers-Daller, 1994: 175)

Whether or not the subject-verb inversion applies is a recurring issue in the discussion of language contact involving ‘verb second’ languages. Treffers-Daller (1994: 176-89) examines the phenomenon in her Brussels data and in other studies of Dutch and German as minority languages. The general trend appears to be that in clauses where the ML is German or Dutch, pre-posed adverbs and adverbial PPs from non-‘verb second’ languages (English and French) do not trigger subject-verb inversion. Sentence-initial adverbs that do not trigger ‘verb second’ are said to be syntactically outside the finite clause. Treffers-Daller calls them ‘hanging adverbs’, or, in GB terms, ‘adjoined to CP’ (1994: 182). Her syntactic analysis is explicated in section 3.3.3 below.

At first glance it is tempting to attribute the syntactic properties of foreign adverbs in German and Dutch to the adverb’s source language. After all, there is no subject-verb inversion in English and French. However, several factors complicate this hypothesis. Firstly, in the monolingual varieties of German and Dutch, the placement rules for pragmatically and semantically similar adverbs are variable. Many native adverbs optionally occur in this initial, ‘hanging’ position where they do not trigger ‘verb second’. De Vriendt & Van de Craen (1993) demonstrate that the placement of various Dutch sentence adverbs is very sensitive to style. At least concerning the adverb *dus* ‘so’, the ‘hanging’ position is by far the most frequent one in spoken language.⁴

⁴ In this connection it is functional to compare sentence adverbs with time and place adverbs. When the latter are fronted, they always occupy the Topic position immediately preceding the finite verb, rather than the position ‘outside’ the finite clause. CS examples with embedded time and place adverbs are few, but so far they corroborate the assumption that the ‘hanging’ position in Dutch clauses is typical of sentence adverbs, and not of foreign adverbs generally. Nortier’s (1990) dissertation cites two examples of adverbs in Topic position in Moroccan Arabic/Dutch (more examples are cited in Chapter 9):

- (1) *temma krijgen ze geld*
 there receive they money
 “There they get money.” MA/Dutch (Nortier, 1990: 152)

Secondly, the ‘contact’ varieties may be subject to change, as a result of which the ‘verb second’ rule applies less consistently than in the standard languages. This is doubtless the case with some linguistic varieties spoken in Dutch and German immigrant communities. However, some caution is called for vis-à-vis such claims, since here again there is a tendency to overlook the structural differences between spoken discourse and the written styles that underlie most grammatical descriptions.

To conclude, there is a strong tendency for sentence-initial French and English sentence adverbs not to cause subject-verb inversion in the contact varieties of German and Dutch. However, it is unclear whether this tendency results from French or English syntactic rules associated with these adverbs, or from the particular pragmatic and stylistic functions these adverbs serve.

3.2.2.2 Swedish/English

Similarly Hasselmo (1972: 175; 1974: 223-4) makes the observation that English adverbs in Swedish/English tend to occur “before the finite verb in the main clause”, that is, between the Subject and the finite verb. In Swedish, like in German and Dutch, the finite verb occupies the second position in main clauses.

- (7) *han finally gick hem*
 he finally went home
 “He finally went home.”
 Swedish/English constructed? (Hasselmo, 1972: 175)
 cf. Swedish: *Han gick äntligen hem. / Äntligen gick han hem.*

The English word *finally* in this example can be interpreted as a marker of subjective modality (‘he should have gone earlier’), as a sequential adverb structuring the narrative (‘to conclude the story, ..’) or as a marker of time (‘after having done x, he went home’). This case represents a more clear-cut violation of the ‘verb second’ rule than the sentence-initial adverbs in CS with German or Dutch. In the latter case it is possible to analyse the foreign adverbs as syntactically outside the finite clause, which is a common position for many native adverbs as well.

3.2.2.3 Irish/English

Further evidence that some embedded English adverbs follow English syntactic rules appears in Stenson’s 1990 article on Irish/English. Stenson (1990: 173, 182-3) notes that a number of English adverbs expressing subjective modality and/or aspect like *still*, *already* and *definitely* tend to occur in ‘medial’ position when embedded in Irish

-
- (2) *gedda kan ik niet*
 tomorrow can I not
 “Tomorrow I can’t [come/go].” MA/Dutch (Nortier, 1990: 128)

clauses, whereas their Irish counterparts are far more common in clause-final position, cf. (8).

- (8) beidh mé thart *definitely* anocht
 be-FUT I around definitely tonight
 “I’ll be around definitely tonight.” Irish/English (Stenson, 1990: 173)

Stenson’s observation particularly concerns adverbs that modify verb phrases or sentences, and express the speaker’s attitude toward what is being spoken of. Their communicative function links these adverbs to the expletive adjectives *fuckin’ friggin’* and *bloody* discussed in Chapter 2. Though the latter are formally attributive adjectives, they do not modify nouns but function as adverbials on the discourse level. It should be remembered that, as EL forms, these expletive adjectives display English prenominal word order (cf. p. 101).

3.2.2.4 Tagalog/English

Tagalog has a number of enclitic adverbials that “with certain storable exceptions, follow the first word in the construction of which they form an immediate part” (Schachter & Otnes, 1972: 411). With gross simplification, these adverbs can be said to occupy the second position in the clause. The details of the placement of various Tagalog adverbs are much more complicated than this. The major modifications of the second position rule are a) monosyllabic pronouns precede any enclitic particle, b) a number of function words either obligatorily or optionally do not occur as “pre-clitics”, so that these are disregarded as potential first words, and c) some adverbs also occur in non-initial sentence constituents in certain syntactic constructions (see Schachter & Otnes, 1972: 411-36).

In a number of Tagalog/English examples cited in Bautista (1980) and Sobolewski (1982), the placement of these enclitic adverbs is probably best explained by Tagalog grammar although they occur in English finite clauses. Consider the following examples (*daw* in (9) and (10) is an epistemic marker; *ho* in (12) marks respect for the addressee).

- (9) (sabi niya,) *each mark* daw *means* (na may *secret admirer* ako)
 EPISTEMIC
 “It is said that each mark means that I have a secret admirer.”
 Tagalog/English (Sobolewski, 1982: 41)

- (10) *beer* daw *is considered a beverage*
 EPISTEMIC
 “Beer is considered a beverage [i.e. non-alcoholic drink].”
 Tagalog/English (Bautista, 1980: 45)

Though examples like (9)-(13) are found in several data sets, Sobolewski (1982: 41) states: “The most common syntactic location of Tagalog adverbials that are used in English clauses and sentences in my data is at the end of an English clause or sentence.” Unfortunately, he does not discuss to what extent this order is compatible with Tagalog grammar.

3.2.2.5 Tidore/Malay

⁵ Concerning this language pair, see Ch. 2, p. 95, and note 17 on p. 95 for Van Staden's definition of the ML.

verb, while its Tidore equivalent *moju* follows it. Likewise, in the first line of (15) we see that the Malay modal *mau* “to want” precedes the Tidore verb *waje* “to say”.

- (14) *tapi nene-re masi so-ninga-ninga moju, ua?*
 but grandmother-this still CAUS-REDUPLIC-‘heart’ still NEG
 “But granny still remembers, right?” Tidore/Malay (Van Staden, 1997)
- (15) *bahasa indonesia-ge waro ma fangato mau waje soma bahasa tidore*
 language indonesia-that know but 1SG·M·SUBJ want say with language
 Tidore
- supaya lebi jelas ua*
 so·that more clear NEG
 “I know [the word] in Indonesian, but I want to say [it] in Tidore so that it is clearer, right?” (Van Staden, forthcoming)

The Tidore/Malay case differs from the previously discussed cases of EL adverbs that defy ML word order rules because this time EL adverbs share this feature with EL modal verbs. The Tidore modals are classified as verbs rather than adverbs because they can stand on their own as main predicates. However, this criterion does not seem to apply to their Malay counterparts: “The NMM [North Moluccan Malay] modals cannot, as a rule, be used as main predicates” (Van Staden, forthcoming, n.10).

3.2.2.6 *Adverbs: summary*

From the above examples it appears that the ML as defined on the finite clause level does not necessarily impose its word order or selectional criteria on sequencing, modal, and aspectual adverbs and, in the Tidore case, modal verbs. Rather, these elements seem to have their own placement rules that apply independently of the language of the rest of the clause, so that the idea of embedding is less applicable here.

The Tagalog adverbial clitics illustrate the point most clearly. The arguments in favour of identifying English as the ML in the cited examples are strong: English grammar determines the internal structure and the word order of the clause, and it even determines which is the first constituent to which the Tagalog adverb cliticizes. Yet the Tagalog adverbial clitic cannot be considered an embedded morpheme if it does not occupy an English slot in the English matrix clause. (In certain instances, such an adverbial slot in the English matrix might be argued for, but not on a consistent basis.)

The idea that modal and sequencing (and perhaps aspectual) adverbs tend to follow their source language syntax, even when they occur within clauses of another language, receives further support from the observations on English adverbs in Irish and American Swedish. In addition, Malay aspectual adverbs and modal verbs in Tidore display the same phenomenon.

Sentence-initial adverbs in CS with the ‘verb second’ languages German and Dutch have received relatively much attention in the literature. However, no definite conclusions can be drawn from the data so far. In CS, English and French sentence adverbs precede German or Dutch finite clauses without triggering subject-verb inversion. Some adverbs also follow the finite verb in a German or Dutch clause. These are the most common positions for native sentence adverbs as well, especially in spoken discourse. Further research and perhaps a more refined classification of sentence adverbs are needed in order to ascertain whether the distribution of English and French sentence adverbs really distinguishes them from their German and Dutch counterparts in codeswitching contexts.

3.2.3 Conjunctions

The use of conjunctions from one language in the context of another is a recurrent characteristic of bilingual conversations. Conjunctions from the ‘other’ language are particularly favoured as discourse markers, perhaps because they are more salient (De Rooij, 1996). Hasselmo for instance notes that the English *and* in American Swedish functions in the first place as a sequencing signal (‘sekvenssignal’, 1974: 231). Likewise, Roy (1979, cited in Lefebvre 1984: 25) observes that the variety of Acadian French spoken in Moncton, Canada, has ‘borrowed’ the English conjunctions *but*, *so*, and *and* which function either as coordinate conjunctions between sentences, or as sequencing devices in the discourse. That is, they do not conjoin French phrasal constituents.

As noted above, the possibility of EL conjunctions, whether subordinating or coordinating, conjoining two ML clauses challenges the classical Matrix Language models because, as function morphemes, conjunctions are not supposed to occur there. This can be solved by re-defining the class of function morphemes that are associated with the ML and analysing the EL conjunctions as embedded content words, or even as EL constituents.

In the latter case, the Monolingual Structure Approach would predict that the distribution of EL conjunctions be governed by ML grammar. Conversely, if the distribution of the conjunctions points to the grammar of their ‘source’ language, rather than to the ML on the finite clause level, it is not really possible to consider them as embedded elements.

3.2.3.1 Shaba Swahili/French

De Rooij demonstrates that in Shaba (or Copperbelt) Swahili/French, the French complementizer *que* has the distributional properties of its Swahili counterpart *asema*.

His main argument is this: unlike *asema* in Swahili, *que* in French does not introduce a clause in ‘direct speech’. More formally, co-referential arguments in the main clause and the subordinate clause introduced by *que* have to agree in person, and there must also be agreement in tense between both clauses. Therefore the French sentences (16)c and d are ungrammatical.

- (16) a. il_i disait toujours: je_i suis maudit (direct speech)
 b. il_i disait toujours qu' $'il_i$ était maudit (indirect speech)
 c. $*il_i$ disait toujours que j_i étais maudit
 d. $*il_i$ disait toujours qu' $'il_i$ est maudit (De Rooij, 1996: 163)
 “He always said: I’m cursed.” / “He always said that he was cursed.”

However, when *que* is used in Swahili contexts, it CAN introduce clauses in direct speech:

- (17) $mungu_i$ a_i -na-sema *que* eh? mu -ntu mi_i -na-mu-fwa(nya) *à l'image*
 God he-TMA-say COMP CL1-man I-TMA-him-make in DET-image

 $yangu_i$
 POSS·1SG
 “God said: I create man in my image.”
 Swahili/French (De Rooij, 1996: 165)

Mungu “God” in (17) is co-referential both with the third person singular verb *anasema* “he says” in the main clause and with the first person singular verb *minamufwa(nya)* “I create” in the complement clause.

According to the author, French *que* does not introduce subordinate clauses in Shaba Swahili, and it is avoided in contexts that require the subjunctive mood in Swahili (De Rooij, 1996: 165-6). Hence *que* in Shaba Swahili/French displays Swahili, that is ML, distribution and can be considered an EL form. (See also Kamwangamalu (1989) and Myers-Scotton (1993b:132) on French *que* in Lingala, another Bantu language of Congo-Kinshasa, former Zaire.)

3.2.3.2 Foreign subordinate conjunctions in German and Dutch

One way to test whether EL conjunctions are syntactically embedded, is to look at matrix languages in which conjunctions are associated with (or ‘trigger’) a particular word order. The West Germanic languages Dutch and German, for instance, have clearly distinct word orders for subordinate and main clauses: The ‘verb second’ rule applies in main clauses (the finite verb occupies the second position, see section 2.2.1 above) and subordinate clauses are verb-final (finite and infinite verbal forms cluster at the end of the clause). Now the question remains whether subordinate conjunctions from languages without a different word order for main and subordinate clauses trigger subordinate clause order in Dutch and German.

Treffers-Daller’s study of Dutch/French contact in Brussels offers a review of some relevant data on Dutch and German in contact with English and French (1994: 189-96). Foreign subordinate conjunctions that fail to trigger verb-final word order are found in all cases. Two examples are reproduced here:

- (18) *tandis que* hier zijn 't stenen
 while here are it stones
 "Whereas it is stone here."
 Brussels Dutch/French (Treffers-Daller, 1994: 190)
- (19) un noh isch de kleinmann nunter, *parce que* ich hab mi dort muen
 and now is the Kleinmann below because I have myself there must
 melde
 register
 "And now Kleinmann is downstairs, because I've had to register myself there."
 Alsatian German/French (Gardner-Chloros, 1991: 169)

Treffers-Daller makes no mention of foreign conjunctions that do trigger verb-final word order, so the overall picture is that foreign subordinate conjunctions differ syntactically from their ML counterparts and do not behave as 'proper' embedded elements in these languages.

In the same vein, Van Ness (1994: 294) postulates that in Pennsylvania German, English *because* 'transfers coordinate clause [i.e., main clause] word order', see (20).

- (20) aβə ɛə hɔd not nɛd g'mɔlgə because əs I s
 nɔx ʃiə tsə fri
 but he has ?⁶ not milked because it is still almost too early
 "He hasn't milked yet, because it's still almost too early."
 Pennsylvania German/English (Van Ness, 1994: 294)

Some caution is called for, however. The observation that EL conjunctions do not trigger verb-final order in Dutch and German faces approximately the same complications that were pointed out with regard to the 'verb-second' rule (section 2.1.2 above).

Firstly, the varieties of Dutch and German that are in close contact with French and English may have undergone syntactic change. This may be particularly the case concerning the immigrant communities in Australia and North America. Thus Enniger & Raith (1988: 287) claim that in the dialect of the Old Order Amish in Kent County, Delaware, German *weil* is followed by main clause order as a result of convergence toward the English norm (see also many publications by Clyne on German and Dutch in Australia, e.g. 1971, 1972). Van Ness (1994: 294), on the other hand, asserts that this is not so in the case of her Pennsylvania German informants.

But more importantly, how do we know that the EL conjunction occupies the slot of an ML conjunction that triggers verb-final order? The latter problem is all the more

⁶ The glosses and translation are mine. The typeface used in the example in Van Ness's article suggests that *not* is an English word, but on the basis of her German translation we would expect a word that means "yet", perhaps German *nɔx*.

intricate if we take into account that monolingual German and Dutch have a number of words traditionally classified as subordinate conjunctions that are followed by either subordinate clause (verb final) or main clause (verb second) word order. The latter order is stigmatized, if not neglected, in prescriptive grammars (Günthner, 1993). Contrary to popular belief, the word order is conditioned by the pragmatic function of the conjunction rather than by dialectal or idiolectal variation. Günthner (1993) demonstrates this for German *weil* “because”: when factual causality is emphasized the verb-final order is required, however, *weil* has a number of other pragmatic functions that require the ‘verb second’ order. Among the uses of *weil* that require main clause order Günthner mentions the justification of a speech act (‘Sprechakt-qualifikation’), see the discussion of example (1) above, and so-called epistemic substantiation (‘epistemische Begründung’). In the latter case the conjunction marks the cause of a belief or a conclusion.⁷ This brings to mind the three kinds of causal relations distinguished in Schifffrin’s (1987: 191-227) analysis of English (*be*)*cause* and *so*: Cause-Result, Warrant-Inference, and Motive-Action. Of these three only the first one is marked by verb-final order in German and Dutch.

My impression is that in Dutch, factual causality on the one hand, and motivation and warranty on the other hand, are more consistently marked by different conjunctions, viz. *omdat* and *want* respectively (cf. German *weil* and *denn / weil*). So in the domain of causal relations, there is less confusion about the appropriate word order in Dutch than there is in German. But other Dutch conjunctions are similar to German *weil* in that different uses require different word orders. For instance the use of *terwijl* “while”, the Dutch counterpart of *tandis que* in (18), as a marker of contradiction seems more compatible with main clause word order than its use as a marker of simultaneity.

So it is possible that foreign conjunctions do not trigger verb-final in Dutch and German matrix clauses simply because they are used in functions that require main clause order. To my knowledge this explanation has not been investigated, perhaps because German and Dutch subordinate conjunctions are presumed to be followed

⁷ Günthner (1993: 42) illustrates the use of German *weil* to mark epistemic substantiation:

- (1) der Bildschirm ist kaputt -
 WEIL - da ist nur noch schwarz aufm Schirm
 “The screen is broken. Because: there’s just black on the screen.”

Here, *weil* introduces the reason why the speaker thinks the screen is broken: because it is all black. Note that this is the reverse of the ‘real world’ causality where the screen is all black because it is broken.

by subordinate clause word order.⁸ It is hardly feasible to examine this hypothesis in relation to the published examples because these are for the most part presented out of context. We do not know, for instance, whether *parce que* in (19) or *because* in (20) marks factual causality or epistemic substantiation.

3.2.3.3 Dutch subordinate conjunctions in other languages

Having concluded that foreign conjunctions do not seem to trigger subordinate clause word order in German or Dutch, we shall want to investigate whether Dutch or German subordinate conjunctions have any syntactic consequences when embedded in other languages. It would be rather surprising if Dutch or German subordinate conjunctions were found to trigger verb-final word order in a language where this word order is not available. I did not come across any example of a Dutch or German conjunction that has a structural effect on a subordinate clause in another language.

Bolle (1995: 292), for instance, mentions that Dutch subordinate conjunctions in Sranan Tongo, a Surinamese Creole language, are followed by normal Sranan SVO order. See the Dutch causal conjunction *omdat* “because” in (21).

- (21) *omdat* mi na wan *suikerpatient*
 because I am a diabetes-patient
 “Because I’m a diabetes patient.” Sranan Tongo/Dutch (Bolle, 1995: 293)

Voigt (1994) gives examples of Moluccan Malay/Dutch where the Dutch conditional marker *als* is followed by a clause that is structurally Malay. In Dutch, *als* triggers a subordinate clause marked by verb-final word order. In the following two conditional clauses, *als dansprogramma’s* in (22) and *als al hari minggu* in (23), the Malay structure is apparent from the absence of an existential marker or copula. In (22) the Malay existential *ada*, which is optional in the spoken language, is omitted.

- (22) *als dans-programma’s* aku lihat *actualiteit* *journaal* *brandpunt*
 if dance-programmes I watch topicality news Brandpunt
televizier
 Televizier
 “If there’s a programme on dance, I watch current affairs programmes, the news, Brandpunt, Televizier [titles of such programmes on Dutch TV]”.
 Moluccan Malay/Dutch (Voigt, 1994: 35)

⁸ Treffers-Daller (1994: 192) does recognize the occurrence of Dutch *omdat* and German *weil* “because” plus main clause word order. According to her, this word order variation is indicative of language change. Günthner (1993), uncovering the pragmatic and syntactic factors that require either word order, convincingly opposes this popular view.

- (23) bangun *altijd* susah *als al* hari minggu
 get-up always difficult if already day Sunday
 “Getting up is always hard if it’s already Sunday.”
 Moluccan Malay/Dutch (Voigt, 1994: 36)

3.2.3.4 Malay conjunctions in Tidore

Tidore (see section 2.2.5 above) has clause-final conjunctions, while conjunctions in Malay are clause-initial. Van Staden (1997) shows that Malay conjunctions precede the clause when used in Tidore discourse, as they do in their source language. Consider example (24) below: *la* is the Tidore clause-final conjunction “in order to”, and *supaya* is its clause-initial Malay equivalent.

- (24) ona gahi ena la, supaya una-ge wo-koliho yali
 3PL do 3SG·N in·order·to in·order·to 3SG·M-that 3SG·M·SUB-returnagain
 “They do that, so that he will come back again.”
 Tidore/Malay (Van Staden, 1997)

The Malay conjunctions in Tidore are perhaps the clearest examples of discourse markers that retain their source language word order properties.

3.2.3.5 Spanish conjunctions in Quechua and Media Lengua

Media Lengua is a contact language spoken in the Ecuadorean Andes, either as a mother tongue or as a second language. According to Muysken’s (1981) analysis, it can be characterised as grammatically a variety of Quechua with a 90% Spanish lexicon. The Spanish lexical items in Media Lengua display the syntactic properties of the Quechua items they replace. This process of massive lexical replacement has become known as relexification. Note that structurally relexification is similar to the process of content morpheme insertion described in Chapter 2. Muysken reports that the Spanish conjunctions occurring in Media Lengua cannot be viewed as cases of relexification:

Are the Spanish conjunctions that we find in Media Lengua cases of relexification of Quechua categories or cases of introduction from Spanish? The latter seems correct because: 1) in Media Lengua the conjunctions are used as in Spanish; 2) the conjunctions coexist with the Quechua cliticized conjunctions; and 3) in Quechua itself we find frequent borrowing of Spanish conjunctions. (...) Thus the introduction of Spanish conjunctions into Media Lengua is the one exception to the idea that Media Lengua lexical categories, arising through relexification, correspond strictly to Quechua categories. This exception may be explained by the fact that conjunctions which occur at discourse level are less closely integrated into the grammar of the language and can be borrowed more easily (Muysken 1981: 65)

The Spanish conjunctions in Media Lengua as well as in (varieties of) Quechua are further examples of foreign conjunctions that retain their source language syntactic properties. Naturally, Media Lengua, being a more or less stable language variety, cannot be equated with CS between two distinguishable languages. But the present state of affairs in Media Lengua results from a bilingual situation where CS is likely to have been widespread, and the product of relexification is structurally parallel to massive content morpheme insertion in CS. Moreover, most of the present day speakers of Media Lengua also speak Spanish, so perhaps it is not even possible to distinguish Media Lengua proper from Media Lengua/Spanish CS.

3.2.3.6 *Conjunctions: Summary*

In some CS contexts such as Swahili/French, Sranan/Dutch and Malay/Dutch, singly occurring ‘foreign’ conjunctions display distributional characteristics of the language of the context, thus suggesting that they are embedded and syntactically integrated in the ML. (The provisional term ‘foreign’ is used here to avoid the term EL.)

In other cases ‘foreign’ conjunctions display distributional characteristics of their source language. The Malay conjunctions in Tidore and Spanish conjunctions in Media Lengua and Quechua demonstrate this most clearly. In such cases one may conjecture that the ‘foreign’ conjunctions impose their source language syntax. Perhaps the French and English subordinate conjunctions that do not trigger verb-final order in German and Dutch constitute examples of the same phenomenon. The evidence from German and Dutch as ML is rather weak, however, so long as it remains uncertain whether the ‘foreign’ conjunctions really occur in a context that requires verb-final order.

However, a generalisation can be made for all the situations examined above: There is no indication at all that the ‘foreign’ conjunction exerts syntactic influence either in the form of word order or by triggering subordinate verb forms. In this respect foreign subordinate conjunctions that have no syntactic consequences range with native subordinate conjunctions in some languages. In other languages such conjunctions differ from native ones while matching with native coordinate conjunctions or other discourse markers. In other words, ‘foreign’ conjunctions are not found in contexts that show a clear hierarchical relationship between the conjunction and the adjacent clause. Consequently it is difficult to establish whether ‘foreign’ conjunctions should be considered as embedded elements.

3.2.4 **Clause-external foregrounding strategies**

The term ‘clause-external foregrounding strategies’ is taken from Foley & Van Valin (1984) who use it to designate topicalisation, left-dislocation and cleft constructions, as against clause-internal foregrounding and backgrounding constructions, viz. passive, antipassive and dative shift. Clause-internal constructions can have both foregrounding and backgrounding functions, e.g. foregrounding passives foreground non-actors, and backgrounding passives background actors. Clause-external

constructions, on the other hand, are exclusively foregrounding. The clause-external constructions considered here are topicalisation and left-dislocation. The difference between these two is that “in left-dislocations but not in topicalizations there is a pronoun in the clause which refers to the clause-initial NP” (Foley & Van Valin, 1985: 300).

Below I will discuss Topic-Comment structures in Japanese/English and so-called discourse emphatic pronouns in CS with Arabic. In Arabic, clause-initial emphatic pronouns are similar to left-dislocated NPs. The Japanese Topic-Comment structure differs somewhat from classical topicalisation, because in the latter case the topicalised constituent leaves a gap in the ensuing clause while this is not necessarily so in Japanese Topic-Comment.

While these foregrounding strategies are difficult to describe in formal syntactic terms, they are pertinent to the syntax of CS. As the term already suggests, clause-external constructions are not explained by the ML on the finite clause level. This becomes apparent when the two languages involved in CS differ in terms of foregrounding constructions and the attested construction must be attributed to one language while the other language is identified as the ML on the clause level. In this respect emphatic pronouns in CS with Arabic and Topic-Comment in Japanese/English illustrate how discourse grammar constitutes a challenge to the study of CS.

3.2.4.1 Topic-Comment structures in Japanese/English

The Japanese Topic-Comment structure is recognisably present in the bilingual speech of second generation Japanese in Toronto, Canada, as described by Nishimura (1986, 1989, 1995a,b; see Ch. 2 section 4.2.1). Topics are fronted and marked by the particle *wa*. This particle may be omitted in spoken discourse. The Topic-Comment structure serves to re-introduce a topic that was mentioned earlier in the conversation, or to mark a contrast between the preceding and the newly introduced topic. Thus, *John wa* could either mean “speaking of John, ..” or “John, as opposed to others”. Japanese Topics can be NPs, PPs, or adverbs. Nishimura, citing Japanese linguists, stresses that no syntactic relationship is necessary between the Topic and the Comment, the only requirement being pragmatic relevance (Nishimura, 1989: 366-7, in a revision of her 1986 point of view). The next two examples show that the Topic constituent does not necessarily have a grammatical function in the following clause (cf. also (28)).

- (25) are wa, zettai ni amerika ga warui
 that TOP absolutely America NOM wrong:is
 “Speaking of that matter, absolutely, America is to blame.”
 Japanese (Kuno, 1973: 253, cited in Nishimura, 1989: 367)
- (26) basho wa, okunai-setsu ga attooteki datta
 place TOP indoor-theory NOM predominant was
 “Speaking of the place [of the murder], the ‘indoor’ theory was predominant.”
 Japanese (Kuno, 1978: 136, cited in Nishimura, 1989: 367)

(27) *Powell street wa we used to call it Little Tokyo*
TOP

(28) kore wa *she was at home*
this TOP

(29) yoroi kabuto o kiteiru yoo ni wa *you can't go*
armor helmet ACC wear way P TOP

(30) *she-wa, took her a month to come home* yo
 she-TOP you know

In all these examples the finite verb is English and therefore the Monolingual Structure Approach identifies English as the ML for the finite clause. The pre-clausal Topic is clearly not embedded in the English matrix structure, nor does it appear to affect its integrity. Hence the Japanese Topics are not outward counter-examples to the MSA, but neither are they explained by it. Instead, Topic-Comment seems to be a higher order matrix structure in which the English finite clause is a constituent. This line of thought is explored later on in this chapter.

⁹ “Topics are realized only in Japanese not in English” (1989: 376). However, her data contain some examples where the Topic is in English except for the marker *wa*, cf. (27) and (30).

Introduction

Arabic makes use of so-called discourse emphatic pronouns. These are free form personal pronouns that commonly precede the clause. However, discourse emphatic pronouns are also found in clause-final or medial position. They mark emphasis in a general sense, but their main function is to signal a change in Topic, sometimes implying contrast, e.g. between YOU and I. The emphatic pronoun has an overt co-referent in the adjacent clause or sentence. The co-referent is either Subject of a finite verb, as marked by agreement, or a resumptive pronoun or, more commonly, a pronominal suffix. The pronoun's co-referent is not restricted to any particular grammatical function in the clause, although, as Topic, the emphatic pronoun is typically co-indexed with the Subject.

Sentence-initial emphatic pronouns can be considered as special instances of the general mechanism of NP left-dislocation that characterises Arabic sentence patterns. However, in the case of Topic pronouns which are co-indexed with the Subject, their comparison with left-dislocation depends on the syntactic analysis of full Subject NPs in Arabic. In most varieties of Arabic both the word orders Subject-Verb and Verb-Subject occur. It is possible to consider the VS order as the basic order from which the SV is derived by means of left-dislocation of the Subject NP. According to this analysis the resumptive pronoun co-indexed with the Subject NP is subsumed under the Subject agreement feature of the verb. The alternative analysis considers both word orders SV and VS as basic, which implies that it is impossible to extract the Subject NP from the clause into the left-dislocated position. The preferred analysis is ultimately a matter of personal taste but it is noteworthy that the SV as opposed to the VS order involves the foregrounding of the Subject NP. Caubet (1993) describes this in detail for Moroccan Arabic; in the following citation she summarises the pragmatic functions of the SV order in clauses where the Subject (referred to as C_0) is a full NP.

En situation d'interlocution, il s'agit essentiellement de valeurs de mise en avant du C_0 ou de comparaison-contraste.

Dans le récit (le conte), il y a une opposition entre, d'une part, une succession d'événements repérés les uns par rapport aux autres, où le verbe est en tête (...), et d'autre part, la description d'un état permanent ou passager, où c'est le C_0 qui est mis en avant (Caubet, 1993, II: 5)

The foregrounding ('mise en avant') function of the SV order associates pre-verbal full Subject NPs with left-dislocated NPs on the one hand and clause-initial Topic pronouns on the other. Indeed, the pragmatic functions Caubet describes for the SV order with full Subject NPs and with Topic pronouns are largely the same (1993, II: 4-9).¹⁰

¹⁰ Caubet notes that in addition to the foregrounding functions shared with full NPs, Topic pronouns which are co-referential with the Subject serve to signal concomi-

Topic pronouns differ from left-dislocated NPs in that they appear as free morphemes (full form pronouns) when dislocated while they are mostly affixes (inflectional morphology) in their ‘normal’ position. In the absence of overt case marking other NPs have the same form whether or not they are dislocated. This difference is not directly related to left-dislocation but follows from the morphology of Arabic pronouns. Obviously, as dislocated constituents, discourse emphatic pronouns cannot be co-indexed with a non-pronominal NP.¹¹ Also, unlike other NPs, discourse-emphatic pronouns cannot be co-referential with full pronouns in the immediately adjacent clause. This fact is explained by a general rule that disallows juxtaposition of identical constituents (cf. Eid, 1991: 46). In conclusion, on the basis of syntactic and pragmatic similarity, I will treat clause-initial emphatic pronouns as special types of left-dislocation, on the assumption that the SV word order in Arabic represents the left-dislocation of Subject NPs.

Before I address emphatic pronouns in CS with Arabic, a few examples will illustrate the use of such pronouns in Moroccan Arabic. Consider the following sequence. (Each time both the emphatic pronoun and its co-referent will appear in bold-face type.)

- (31) **bqi-na** **mša** **bašt-na** **b** **kunṭak**
 remain-1PL with RECIPROCAL-1PL with contact

huwa **y-gul** **had** **l-kelma** **hakka**
 3M 3-say DEM DEF-word this-way

ana **n-ḍemm..** **n-ḍemmen**, **n-xemmem** **fi-ha**, **ne-ḥfeḍ** **fi-ha**
 1SG 1-interna.. 1-internalize 1-think in-3F 1-retain in-3F

“We kept on seeing each other. He says this word like this, and I internalize, I think about it, and retain it.” Moroccan Arabic (Mustafa)

In the above fragment, Mustafa talks about how he learned the Berber language while he was in the Netherlands. In the first line, the first person plural refers to the speaker and his Berberophone friend. He then shifts to his friend as the Topic, as indicated by the third person masculine pronoun *huwa*, and back to himself again, as indicated by *ana*. In both cases, the discourse emphatic pronouns are co-referential with the Subject agreement prefix on the following verb.

Now (32) is a straightforward example showing that emphatic pronouns are not necessarily grammatical Subjects. *ana* “I” is co-indexed with the possessive suffix *-i* in *smiyt-i* “my name”.

tance in Moroccan Arabic (1993, II: 9).

¹¹ This feature distinguishes them from the sentence initial pronouns that mark questions in Egyptian Arabic (Eid, 1992b).

- (32) **ana** smiyt-i hayat
 1SG name-1SG Hayat
 “My name is Hayat.” Moroccan Arabic (Hayat)

Note that here *ana* is discourse emphatic because *smiyt-i hayat* suffices to express “My name is Hayat”. In the next example *hna* “we/us” recurs as the object suffix *-na* in the following clause. (This clause may itself be analysed as a left-dislocated object that is co-referential with the object suffix *-hūm*.)

- (33) **hna** lli ka-t-naseb-**na** bezzaf, l-luġa lli ka-t-naseb-**na** **hna-ya**
 1PL REL ASP-3F-suit-1PL much DEF-language REL ASP-3F-suit-1PL 1PL-EMPH
 ġadi n-gul-hūm-l-ek :
 FUT 1-tell-3PL-to-2SG
 “As for us, what suits us very much, the language[s] that is [are] useful to us,
 I’ll tell you which ones: (..)” Moroccan Arabic (Mustafa)

hnaya in (33) is an example of a clause-final emphatic pronoun. It refers back to the object of the preceding verb *ka-t-naseb-na* “it suits us” (the emphatic extension *-ya* is not critical here).

The discourse emphatic pronoun does not preclude the occurrence of another left-dislocated NP. In the next sentence, both the Subject and the object are left-dislocated and recur in the verbal suffixes *-t* and *-ha* respectively. In addition, the adverb *hna* “here” is topicalised in the position preceding the verb. Adverbs and PPs can only be topicalised in Arabic, not dislocated (cf. Lalami, 1996).

- (34) w-**ana**_i š-šelḥa_j **ana-ya**_i, hna, hna tšellem-t_i-**ha**_j f.. eh f
 huḷanda
 and-1SG DEF-Berber 1SG-EMPH here here LEARN-1SG-3F in er in Holland
 “As for me, Berber, I learned it here in Holland.” Moroccan Arabic (Mustafa)

Eid (1985, 1992a, 1996) and Barhoum (1994: 103) maintain that discourse emphatic pronouns are Subjects, and indicate Subject shift. Whether these pronouns are Subject or Topic is crucial for their syntactic analysis in an X-bar framework, as we will see shortly. Setting aside formal considerations concerning X-bar syntax, the language facts are decisively in favour of the Topic hypothesis. There is no independent criterion to distinguish between discourse emphatic pronouns that are co-referential with a following (or preceding) Subject and those whose co-referent has another grammatical function in the clause. Moreover, emphatic pronouns can be perfectly appropriate in contexts where there is a continuation of the grammatical Subject, see (43) and (45) below (more examples in Eid 1996: 13, 14). Consequently, these pronouns cannot mark Subject shift and are better analysed in terms of pragmatic functions.

Discourse-emphatic pronouns in codeswitching

The use of discourse-emphatic pronouns is far more pervasive, and considerably less marked, in Arabic than in English or Dutch, or French. Since Arabic is a so-called ‘pro-drop’ language and non-Subject pronouns are suffixes, most - but certainly not all - free form pronouns are redundant and may therefore be classed as discourse-emphatic. The occurrence of Arabic Topic pronouns in otherwise non-Arabic texts is a characteristic of codeswitching with Arabic. It has been reported for the Maghreb region, Malta, Egypt and the Levant.¹² Examples from various contact situations are presented below.

- (35) *nous, quand on est cassés, on fabrique de l’os,*
 we when INDEFPRON is broken INDEFPRON produce PART DEF-bone
huma, ils n’en fabriquent pas
 3PL they NEG-of-it produce NEG
 “We, if we have something broken, we produce bone tissue, but they [= aged people] don’t produce any.”
 Moroccan Arabic/French (Slaoui, 1986 Annexe 3: 11)

In (35) we find two discourse emphatic pronouns, the French *nous* and Moroccan Arabic *huma*, that signal the disparity of two groups of people. Both are co-indexed with French Subject pronouns, the former with *on*, the latter with *ils*. In (36) the Moroccan Arabic emphatic pronoun *nti-ya* (2F) is co-indexed with the Dutch non-Subject pronoun *jou* “you”, complement of the preposition *voor* “for”. The PP *voor jou* is in its turn topicalised to occupy the first position in the Dutch clause.

- (36) *ʕend-na l-mašakel zeʕma n-gul-u bħal bħal. ik eh ..*
 at-1PL DEF-problem*PL EPIST I-say-PL same same I er
muhiimm nti-ya voor jou was het misschien ehm iets moeilijker
 anyway 2F-EMPH for you was it maybe er somewhat more*difficult
 “We have - let’s say - the same problems. I er .. Anyway for YOU it was maybe somewhat more difficult.” Moroccan Arabic/Dutch (Samir)

The next example shows an interaction between a teacher and students during a home-economics lesson in a Maltese classroom. Note that the two subsequent occurrences of *aħna* (1PL) here are not co-referential. In the teacher’s turn, *aħna* includes the teacher and the students (‘we-inclusive’), and is co-indexed with the ensuing English

¹² In addition to the printed sources cited here, the phenomenon was discussed on the Arabic Linguistics list on Internet (arabic-l@byu.edu, moderator Dil Parkinson) on 12 and 13 March 1996. Under the title “double subjects”, various contributors reported to have heard Arabic full pronouns preceding clauses in English or French in Egypt, the Levant and North Africa.

pronoun *we*; the second *aħna* is a ‘we-exclusive’, denoting the student and her/his family, and is referred to by the 1st person plural suffix *-na*.

- (37) Students imma dawk *flat* għand-hom
 but DEM•PL flat at-3PL
 “But they live in a flat.”
 Teacher **aħna** *we are talking about a flat*
 1PL
 “‘We are talking about a flat.’”
 Student ija **aħna** dar għand-na
 yes 1PL house at-1PL
 “‘Yes, we live in a house.’”
 Maltese/English (Camilleri, 1996: 99)

Concerning eastern dialects of Arabic, Eid (1992a, 1996) and Barhoum (1994) address the same phenomenon in the speech of ‘Egyptian-Americans’ and Americans of Levantine origin respectively. Some of their examples are reproduced here.

- (38) **ʔihna** *it is none of **our** business*
 1PL
 “‘It’s none of OUR business.’”
 Levantine Arabic/English (Barhoum, 1994: 100)
- (39) fa **hiyya** *psychologically **she**’s in Egypt*
 so 3F
 “‘So psychologically she’s in Egypt.’”
 Egyptian Arabic/English (Eid, 1992: 59)
- (40) yaʕni *definitely **ʔana** I picked up a lot of things*
 meaning 1SG
 “‘That is, definitely, I picked up a lot of things.’”
 Egyptian Arabic/English (Eid, 1992: 59)

Despite the example cited here in (38), where the Arabic pronoun is co-referential with the English possessive *our*, Barhoum only speaks of “doubling of first-person subject pronouns” (1994: 103). All the examples he cites indeed concern the first person, and, both in his and Eid’s data, most clause initial Arabic pronouns are co-referential with free form English Subject pronouns, e.g. *she* in (39) and *I* in (40).

Consistent with her 1985 analysis of Arabic clause initial pronouns as Subjects rather than Topics, Eid (1992a, 1996) believes that ‘doubling of Subject pronouns’ is the process involved in examples like (39) and (40) and she coined the term ‘pronoun doubling’ for this characteristic of codeswitching with Arabic. Her 1992 suggestion that the redundant English Subject pronoun is needed to meet the agreement

requirements of verbs in Arabic (1992a: 59-62) is incompatible with the Monolingual Structure Approach. According to the MSA, agreement properties of Arabic verbs play no role in an English finite clause. On the contrary, the English pronoun meets the requirements of the English verb (cf. Myers-Scotton et al., 1996: 32, 36). Seen from the standpoint that Arabic discourse-emphatic pronouns are Topics, ‘Subject pronoun doubling’ is only an impression resulting from the confluence of two independent factors: the strong but incomplete correlation of Topic and Subject in Arabic and the fact that English unlike Arabic is not a pro-drop language so that the English Subject pronoun is obligatory in the following English finite clause.

Whether Topics or Subjects, it is improbable that these pronouns are embedded in a French, Dutch or English sentence structure. Firstly, personal pronouns, being paradigmatically organised function morphemes, are less likely to be EL forms from a cross-linguistic perspective (but see Chapter 2 section 2.2). However, more importantly, this kind of discourse-emphatic pronoun is uncommon especially in English and Dutch. Myers-Scotton, Jake & Okasha remark that, in English, topicality is marked on pronouns in argument position by means of higher pitch, e.g. *So you want to go to the movies, well I [higher pitch] want to go to the opera* (1996: 30). Alternatively, phrases like *as for me, speaking of him* can introduce topics. Dutch is very similar to English in this respect. The profusion of such pronouns in both monolingual Arabic and codeswitching varieties, as well as the fact that the pronouns themselves are invariably in Arabic,¹³ clearly shows that we are dealing with a discourse organisation device that must be attributed to Arabic grammar.

Summary foregrounding strategies

With regard to Topic-Comment in Japanese/English and emphatic pronouns in CS with Arabic, two examples were presented of how information packaging and sequencing strategies on the discourse level interfere with the syntactic analysis of bilingual conversations. While the Topic-Comment structure is recognisably Japanese and the use of discourse emphatic personal pronouns is clearly an Arabic foregrounding mechanism, they can also occur with finite clauses in another language, e.g. English. This suggests that the finite clauses participate as constituents in a higher order matrix structure of the type Topic-Comment. In Arabic, this matrix has more of a syntactic nature because of the requirement for Topic pronouns to have a co-referent in the adjacent clause, just like in (other types of) left-dislocation.

¹³ Jake (1995: 279-80) and Myers-Scotton et al. (1996: 27-30) discuss Bentahila & Davies’s (1983: 313) MA/French example *moi dxel-t* ‘I went in’ with the French discourse emphatic 1SG pronoun *moi* preceding a finite verb in Arabic (go.in· PERFECT-1SG). However, this is a hypothetical example of a pattern that Bentahila & Davies report does NOT occur in their data, even though it was found “acceptable” by an unspecified panel of judges.

3.2.5 Discourse grammar: recapitulation

Discourse markers, conjunctions, sentence adverbs, and foregrounding strategies all function on the discourse level, either as a means of sequencing sections of the discourse, or as a means to express an attitude toward the interlocutors or toward what is being said. These expressions of discourse grammar also have in common that they cannot entirely be accounted for in terms of clause syntax. In CS contexts this means that they cannot, or at least not consistently, be associated with the Matrix Language on the finite clause level. Hence, they cannot be explained within the embedding paradigm developed so far in Chapter 2.

Some aspects of discourse grammar are unequivocally attributable to the source language of their surface form. The Japanese and Arabic foregrounding constructions discussed above are clear examples of this. The Topic constituents in Japanese/English, which are often in Japanese and/or marked by the particle *wa*, must be interpreted within the Japanese Topic-Comment construction even if the Comment is an entirely English clause. Likewise, the use of Arabic free form personal pronouns preceding clauses in another language can only be accounted for with reference to Arabic discourse grammar. We saw furthermore the particular tendency of sentence (modal) adverbs to display a distribution, both on the syntactic and on the discourse level, that must be attributed to their source language, even if they occur in or adjacent to a clause for which another language is identified as the ML.

In the case of conjunctions and other clause-external discourse markers, their association with either their source language or the language of the adjacent finite clause is more ambiguous. I conjecture that subordinate conjunctions which occur in the context of another language tend to function on the broader discourse level rather than as subordinators in complex sentences. Conjunctions that signal sequences in the discourse structure as well as other clause-external discourse markers often lack syntactic properties that would unequivocally associate them with one of the languages involved in CS. Examination of their distribution in the discourse and the pragmatic functions they express may relate them to the grammar of either language. The available documentation on this matter is scarce and does not solve the question. On the one hand, there is the French complementizer *que* in Shaba Swahili/French which functions as the Swahili (ML) conjunction or discourse marker *asema*. On the other hand, the syntactic distribution of Malay conjunctions in Tidore and of Spanish conjunctions in Media Lengua and Quechua must clearly be attributed to their source language. In the same vein, English *well* is reported to have the syntactic distribution of its source language when used in German dialects of Texas and Indiana. Further study will be required to warrant any generalisation on this matter.

3.3 A syntactic representation of discourse grammar

We saw how discourse markers and foregrounding constructions challenge the MSA. Here, I will discuss how these phenomena can be dealt with in a syntactic approach, in particular X-bar syntax. First I will explain Eid's and Myers-Scotton's proposals to associate the Arabic discourse emphatic pronouns, in particular, with a syntactic

(matrix) structure on a level above that of the finite clause. Analogous analyses are applied to sentence adverbs, discourse markers, conjunctions, and even the Topic-Comment structure in Japanese. Then I will evaluate these proposals (4.2), pointing out some syntactic complications in 4.3 below.

3.3.1 A syntactic analysis in GB tradition

Eid (1996), Jake (1995) and Myers-Scotton, Jake & Okasha (1996) discuss the syntactic position of Arabic discourse emphatic pronouns within a structural configuration called Complement Phrase (CP). (See also Eid, 1992b for a similar account of clause-initial pronouns that mark questions in Egyptian Arabic.) The basic structure of the CP within a Chomskyan framework is given in *Fig. 3.1*.

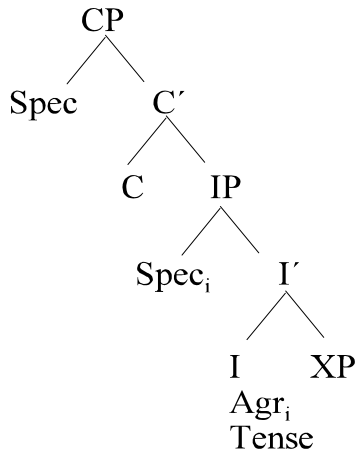


Fig. 3.1. CP structure (Eid, 1996: 16)

In the above structure, C (Complementizer) is head of CP, IP is complement of C and is headed by I (Inflection), and I includes Agreement and Tense. Agreement in I is coindexed with Spec (Specifier) of IP (Eid, 1996: 16). In this representation of the sentence structure, the grammatical function of Subject is realised in the Spec of IP position, while Topic occupies the Spec of CP position (Eid, 1992b: 121).

As for the clause-initial emphatic pronouns in Arabic (see section 2.4.2 above), Eid (1996: 17) locates them in the Spec of IP position. This is in keeping with her assumption that these Arabic pronouns are Subjects (Eid, 1985, 1992a).

On the other hand, Myers-Scotton and her associates perceive these same pronouns as Topics, and accordingly place them in the Specifier position of CP (Jake, 1995; Myers-Scotton, Jake & Okasha, 1996). So, for (41) and (42), two representative examples of the Arabic discourse emphatic pronoun, they propose the structures depicted in *Figures 3.2* and *3.3* respectively.

- (41) **nta tu vas travaill-er**
 2M you go work-INF
 “You are going to work.” MA/French (Bentahila & Davies, 1983: 313)
- (42) **yaʕni ʔana I was really lucky**
 it•means 1SG
 “That is, I was really lucky.” Egyptian Arabic/English (Eid, 1992a: 59)

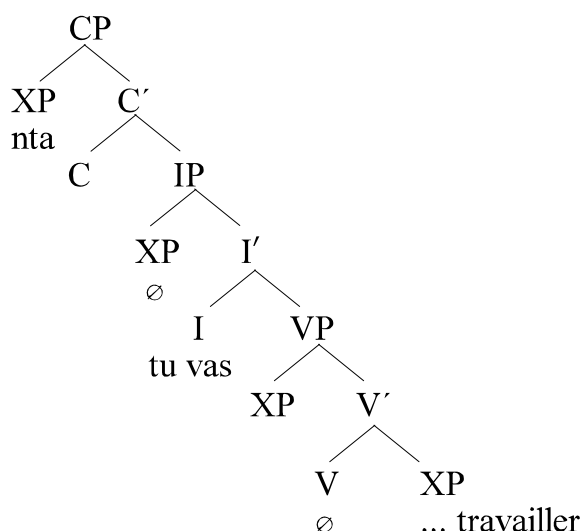


Fig. 3.2. Syntactic analysis of (41).
 (Myers-Scotton, Jake & Okasha, 1996: 28)

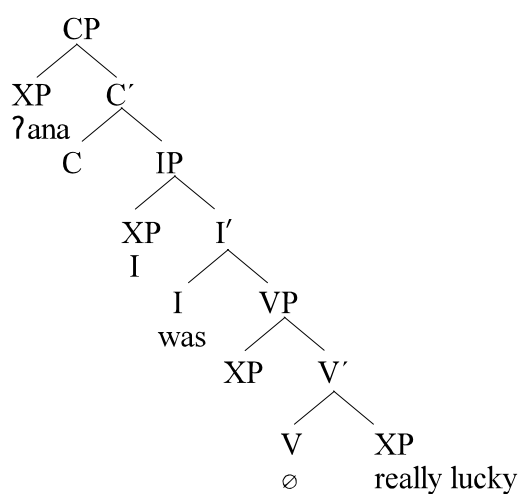


Fig. 3.3. Syntactic analysis of (42).
 (Myers-Scotton, Jake & Okasha, 1996: 29)

In principle Subjects are located in the Spec of IP, but according to Myers-Scotton et al. (1996: 31), French clitic pronouns such as *tu* in (41) are realised as agreement features in the INFL node.

Diverting away from the internal structure of the IP constituent, *Figs. 3.2 and 3.3* show that Myers-Scotton and her associates view the CP as the relevant matrix structure (cf. the discussion of the MLF model in Chapter 1). Where an Arabic emphatic pronoun precedes a finite clause in English, French or Dutch (cf. examples in 3.4.2), they propose an Arabic CP as the matrix structure, and the finite clause (IP) is viewed as an embedded constituent, ‘EL island’ in their terminology.¹⁴ Subsequently the emphatic pronouns are neatly incorporated into a matrix language model. This analysis, which I will call the CP analysis hereafter, explains two features of the emphatic pronouns in CS with Arabic, viz. the occurrence of such pronouns in mixed discourse, and their order relative to the finite clause (IP).

3.3.2 Primary evaluation of the CP analysis

The CP analysis is a meaningful contribution to the development of a model because it associates morphemes that function in the domain of discourse marking with a matrix structure (CP) derived from their source language. This matrix language determines the distribution of these discourse marking/organising morphemes on both the syntactic level and the level of discourse structure. Now this gives the CP analysis a certain appeal since the overall impression is that discourse markers tend to function in accordance with their source language rather than with the language of the finite verb (or the IP). (The French complementizer *que* in Shaba Swahili remains a clear exception, however.)

Many of the other problematic cases involving discourse markers and conjunctions that were treated in sections 3.1 to 3.4 above may be solved in parallel fashion by associating them with a matrix structure above the finite clause level. Sentence adverbs and other discourse markers such as *well* and *so* can be located in Spec CP, for instance, and conjunctions in C. Again, this helps explain the presence of the adverb, discourse marker or conjunction and its order relative to the IP if this order is associated with the source language of the discourse marker. It could explain, for instance, why the English causal conjunction *because* does not trigger verb-final order in Pennsylvania German whereas German *weil* does trigger this order.

In the Monolingual Structure Approach, unlike Myers-Scotton’s MLF model, subsequent levels of ML and embedding are possible. Hence the adoption of the CP as a matrix structure above the finite clause does not disturb the concept of matrix language that has been presented so far. Within a language *x* CP a language *y* IP could

¹⁴ Note, however, that the outer structure containing the Topic pronoun does not serve to identify the ML. Myers-Scotton and her associates insist that the ML is identified by examining a ‘discourse-relevant’ sample (Myers-Scotton et al., 1996: 10 n. 2).

be an inserted constituent and at the same time function as a matrix structure in which lower order constituents are embedded.

While the ‘CP analysis’ thus offers an elegant way out for some intricate problems concerning the status of adverbs, discourse markers and conjunctions in a matrix language model, there are a number of considerations that call for some reserve. These are of two kinds.

The IP as an inserted constituent

Firstly, some objections concern the EL constituent status of the IP. It is somewhat counter-intuitive to associate a single adverb or discourse marker with a matrix structure that encompasses the entire finite clause, and to define the latter as an embedded constituent. The occurrence of conjunctions, adverbs and discourse markers in an ‘other language’ context is very frequent in CS in general. If we adopt the CP analysis for these phenomena, this implies that the insertion of IP constituents is a very common CS mechanism. We may view this as a simple observation, however, it contrasts with the restrictive nature of constituent insertion as observed for NP, PP or VP constituents. Below the IP level at least, it turns out that the more complex the constituent, the less the likelihood of its insertion (see the description of Dutch insertions in Chapters 5-8).

The internal structure of CP

Secondly, the hierarchical structures in *Figs. 3.1-3.3* above look neat and attractive, but the exact syntactic analysis of discourse marking and clause-external foregrounding constructions remains rather complex. This is particularly so when a) there is more than one discourse marker, and b) the discourse marker is neither clause-initial nor clause-final. Generally speaking, it is questionable whether the binary branching hierarchical structures of X-bar syntax can handle the syntactic aspects of discourse organisation. To examine some of these problems, I will first reconsider in some detail the Arabic emphatic pronouns discussed in 3.4.2 and in the explication of Myers-Scotton et al.’s CP structures. Then I will briefly reexamine X-bar analyses that have been proposed for the Japanese Topic-Comment structure, sentence adverbs and conjunctions. I will merely be pointing out the complications, leaving it to the specialists in the field to develop solutions within X-bar syntax.

3.3.3 Complications concerning the CP analysis

3.3.3.1 Discourse emphatic pronouns in Arabic (II)

The analysis of discourse emphatic pronouns in the CP structure faces a number of complications.

1) Emphatic pronouns are also found ‘clause-internally’, immediately following their co-referent which is either a pronominal suffix (verbal or prepositional complement,

possessive marker) or the Subject agreement feature on the finite verb, as illustrated by the next example.

- (43) baš w-ila gal-l-ek ši waḥed mekken-l-i hadi, ka-te-šref
 so•that and-if say-to-2SG INDEF one hand-to-1SG this ASP-2-know
 ula bḡi-ti nta t-sewwel šla ši ḥaža
 or want-2SG 2M 2-ask for INDEF thing
 “So that, if someone says to you ‘hand me this’, you’ll know. Or if YOU want to ask for something.” Moroccan Arabic (Mustafa)

In (43) Mustafa explains that he started to learn Dutch in order to be more successful on the job market. The pronoun *nta* in the second line marks the topic shift from *ši waḥed* “someone” in the first line. Note that *nta* occurs between the finite verb *bḡit* “you want” and the subordinate verb *tsewwel* “you ask”.

I have not yet found CS examples of this. However, we may assume that the CP structure in which Topic is realised in the Spec of CP, as proposed by Jake (1995) and Myers-Scotton et al. (1996), applies to monolingual Arabic as well. Surface forms like (43) above suggest that there are more Topic positions in the CP structure or perhaps that the clause-initial position of Topic pronouns results from some kind of optional movement.

2) According to the structures depicted in *Figs. 3.1-3.3* the Spec of CP position where emphatic pronouns are realised is located on the left-hand side of the C node. In principle, this implies that anything in the C node follows the emphatic pronoun in the linear order. In fact, conjunctions typically precede the emphatic pronoun in the surface structure, as in the next two examples.

- (44) walakin **ana ik** pas overal
 but 1SG I suit everywhere
 “But I [as against my husband], I can live everywhere.” MA/Dutch (Hayat)
- (45) *I found out inn* **ana I** *spend most of my time with my boyfriend*
 that 1SG
 “I found out that I spend most of my time with my boyfriend.”
 Egyptian Arabic/English (Eid, 1996: 13)

It was conceivably this same syntactic complication that led Eid (1985, 1992a, 1996) to interpret the emphatic pronouns as Subjects, which are associated with the IP constituent to the right-hand side of C, rather than as Topics. Eid (1996) examines several alternative analyses that locate the Arabic full pronoun either in the Spec of IP, or together with *inn* “that” in the C node (see *Fig. 3.1* above).

3) The question marker is another element that, according to the X-bar scheme, is located in C. Moroccan Arabic has the optional particle *waš* that introduces yes/no questions. This marker also commonly precedes the Topic pronoun, thus providing another case where the Spec of CP surfaces at the right-hand side of the C position:

- (46) *waš nta-ya mši-t, mši-t ſend-hüm te-ṭleb bdak kursus*
 Q 2M-EMPH go-2M go-2M at-3PL 2-ask for DEM course

ula huma lli gal-u-l-ek : rweḥ, (..) ?

or 3PL REL tell-PL-to-2SG come-IMP

“Did you go to them to ask for that course, or was it them who told you
 ‘Come! (..) ?’ Moroccan Arabic (Samir)

4) As far as I can judge, discourse emphatic pronouns that occur clause-initially correspond formally to left-dislocation with full NPs (cf. the discussion at the beginning of section 2.4.2 above). Regarding left-dislocation in GB syntax, Lalami (1996) shows that Moroccan Arabic left-dislocated NPs are ‘base-generated’, that is, they are not themselves moved constituents. In addition, she argues that they are located in a category that is different from C (or COMP), Spec of CP, and Spec of IP. They are, indeed, “generated higher than CP” (Lalami, 1996: 128).

Left-dislocated NPs must be exiled to a higher node since the Spec of CP position is traditionally reserved for Topic, and a left-dislocated NP can co-occur with a topicalised PP or Adverb in the same sentence. Discourse emphatic personal pronouns can co-occur with topicalised constituents too. Worse still, all three of them can be juxtaposed, as we saw in (34), which is repeated here as (47). *hna* “here” is the topicalised adverb, *š-šelḥa* “Berber” is a left-dislocated NP, and *ana* and *anaya* are two instances of the emphatic pronoun.

- (47) *w-ana š-šelḥa ana-ya, hna, hna tſellem-ti-ha f. eh f huḷanda*
 and-1SG DEF-Berber 1SG-EMPH here here LEARN-1SG-3F in er in Holland
 “As for me, Berber, I learned it here in Holland.”
 Moroccan Arabic (Mustafa)

To conclude, X-bar syntax still has a long way to go in order to explain the word order properties of the emphatic pronouns in Arabic. So far the facts are difficult to handle within a binary branching structure such as the CP. In order to map the attested ‘surface’ structures onto the hierarchically ordered Spec CP, C and IP nodes, various movement rules must be assumed. I am unqualified to judge whether such movement rules are desirable.

3.3.3.2 Topic-Comment structures

With respect to the Japanese Topic-Comment structure (see Ch. 2 section 4.2.1, and section 2.4.1 above), Jake (1995: 286) proposes to locate the Topic constituent in the Spec of CP. In the following example, Jake’s bracketing is reproduced.

- This analysis, which is analogous to her treatment of Arabic discourse emphatic pronouns, obscures the fact that there are important structural differences between these two foregrounding constructions. While the clause-initial emphatic pronouns in Arabic can be viewed as a type of left-dislocation, in which the clause-initial NP has an obligatory co-referent in the ensuing clause, the relationship between Topic and Comment in Japanese is only defined by pragmatic relevance (Nishimura 1989: 367). The Topic-Comment construction cannot be defined in syntactic terms. For this reason Nishimura (1989) abandoned her initial 1986 analysis that was similar to the CP analysis in Jake (1995).

Recall that Treffers-Daller proposes to view sentence-initial adverbs as ‘adjuncts to CP’ if they do not trigger ‘verb second’ in Dutch (see section 2.2.1 above). Her ‘adjunct to CP’ position is higher in the syntactic tree than the Spec of CP node. Consider Treffers-Daller’s representation of the syntactic structure of example (6) above, repeated here as (49).

- (49) *d'ailleurs* 't gasthuis heeft 't ook geconfirmeerd
anyway the hospital has it as well confirmed
“Anyway, the hospital has confirmed it as well.”
Brussels Dutch/French (Trefers-Daller, 1994: 175)

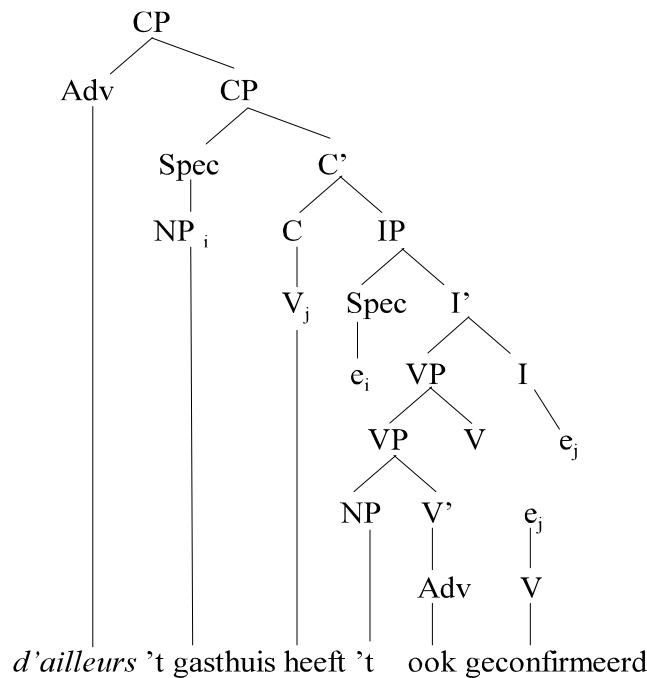


Fig. 3.4. Sentence initial adverbs in a CP structure
(Treffers-Daller, 1994: 182)

Note that, unlike Myers-Scotton and Jake, Treffers-Daller does not explicitly attribute the highest CP structure to either language. However, it appears that the structure in Fig. 3.4 must be Dutch because “French adverbs are adjoined to IP and not to CP” (1994: 182). In this tree, the Subject NP *'t gasthuis* “the hospital” has moved to the Spec CP node, and the finite verb to the C node. What concerns us here is that the adverb *d'ailleurs* takes the left-most position outside (the lower) CP.

According to Treffers-Daller, “Dutch adverbs, PPs and other constituents normally appear in the Specifier position under CP. In that case, the subject remains in its original position” (1994: 182). In other words, Dutch adverbs and other constituents are normally followed by the finite verb. Therefore one reason to assign a position left of Spec CP to the French adverb is that the finite verb is in the third position in (49). Another reason is that French sentence adverbs can occur on the left-hand side of dislocated elements (1994: 181). Left-dislocated NPs, however, are not immediately followed by the finite verb in Dutch, and therefore they cannot be assumed to take the Spec CP position either. This position is, after all, reserved for topicalised constituents. So another node adjoined to CP must be assumed for left-dislocation. In addition, a coordinate conjunction can precede a sentence-initial adverb, which requires an ultimate left-most node. In sum, an elaborate hierarchical structure must be assumed above the CP level, because several discourse-related elements (left-dislocations, conjunctions, adverbs) can be combined at the beginning

of a clause. Since all these markers are optional, the X-bar analysis of most utterances contains various empty nodes.

The picture is further complicated by the fact that the same sequencing adverbs in Dutch main clauses can occur in two other syntactic positions, viz. the Topic position immediately preceding the finite verb, and a position at the right hand side of the finite verb (cf. 3.2.1). As a result, one wonders whether the syntactic properties of sentence adverbs and other discourse markers would not be better explained by other principles, such as pragmatics, or by relating them directly to individual lexical items.

3.3.3.4 *Conjunctions*

One of the motivations for qualifying complementizers and subordinate conjunctions as governors in GB syntax, is that these elements coincide with a subjunctive mood on the verb in the following clause, or, in some languages like Dutch and German, a distinct word order. The subordinate clause is regarded as the complement of C. CS data seem to reinforce this idea to some extent. In German and Dutch, verb-final order does not follow after conjunctions from other languages such as French or English. This might be explained as a property of the French and English conjunctions since there is no subordinate clause word order in these languages. On the other hand, Dutch and German subordinate conjunctions do not lead to verb-final order in languages that do not normally have such word order, as we have seen in 3.3.4 above. Therefore the generalisation can be made that ‘foreign’ conjunctions typically do not have any structural effect on the clause they mark. This is due either to their status as ‘unintegrated’ foreign forms, or to their serving discourse functions that are not structurally marked inside the clause.

In any case, the fact that ‘code-switched’ conjunctions fail to produce structural impact on the clauses they conjoin makes it less appealing to associate them with the C node in the hierarchical structure CP. This is even more the case if the absence of government sets them apart from native conjunctions.

3.3.4 Summary CP analysis

Several aspects of discourse grammar cannot be dealt with satisfactorily within the Monolingual Structure Approach as advanced in Chapter 2. Summarising the examples presented in section 2 above and the discussion of their syntactic representation in 4 below, we can discern two main reasons why the notion of matrix language does not work here.

Firstly, many discourse markers do have syntactic properties but there is no hierarchical relation either way between them and the finite clause. Secondly, there are aspects of discourse organisation that may be described in terms of matrix structures but the syntactic definition of these structures is problematic. The latter case is best exemplified by the Topic-Comment construction in Japanese/English CS. This is a decidedly Japanese matrix structure consisting of two constituents: a

nominal, prepositional or adverbial Topic constituent, optionally marked by the particle *wa*, and a pragmatically related Comment. In Arabic there is more of a syntactic bond between the Topic pronoun and the Comment: The emphatic personal pronoun has a co-referent in the Comment sentence with which it agrees in person, gender and number. In both cases the foregrounding construction can be viewed as a matrix structure in its own right, as pointed out by Nishimura for Japanese and by Myers-Scotton and her co-authors for Arabic. The schematic representation of the Japanese and Arabic foregrounding structures is redrawn in Figs. 3.5 and 3.6. The arches represent the pragmatic structure, leaving the syntactic analysis undecided.

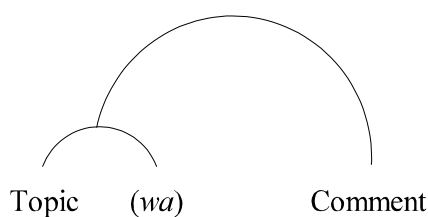


Fig. 3.5. Topic-Comment in Japanese

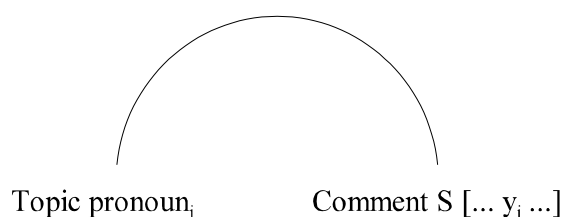


Fig. 3.6. Topic pronouns in Arabic

A syntactic definition that would bring these foregrounding constructions together under one label is not particularly evident. The proposed X-bar analyses express the idea that the foregrounding constructions constitute a higher order matrix structure, but they suggest more syntactic rigour and uniformity than can be justified by the data.

In both cases the pre-clausal Topic is itself the only indication of the existence of the higher order Topic-Comment matrix structure, which is assumed in order to explain the presence of the Topic. For this reason the recognition of these foregrounding constructions as matrices above the finite clause level does not warrant the same type of generalisations as the ML on the finite clause or lower constituent level. Remember that, on the finite clause level, the ML is defined by an independent criterion (the finite verb inflection), it explains the selection and distribution of the other constituents in the clause and applies to different finite clause types (e.g. declarative and interrogative, main and subordinate clauses). The ML on the constituent level covers a number of different constituent types (NP, PP, VP and sub-categories of these) and different types of insertions.

Foregrounding is a sequencing mechanism that must be described in terms of structures, not by pointing out the functions of individual morphemes. Where sequencing (and evaluating) of discourse is marked by specific morphemes, the matrix structure approach is less evident. First it was shown that several kinds of discourse marker (adverbs, conjunctions, particles) are difficult to include in a matrix language model of CS. They cannot very well be considered as embedded elements because the matrix structure defined as the finite clause does not systematically impose its ML distributional properties on discourse markers from the other language. In this

respect, such elements differ from argument constituents, whose distribution in the clause is determined by the language of the finite verb.

Because of their peripheral syntactic status in the clause, it is often difficult to ascribe the distribution of conjunctions and discourse markers to either language involved in CS, unless they mark functions that are marked in only one of the two languages. In other cases, discourse markers, especially conjunctions, seem to follow ML distribution (e.g. Shaba Swahili/French, Moluccan Malay/Dutch). Sentence adverbs, on the other hand, display a tendency to maintain their source language word order when embedded in ML finite clauses. However, no clear generalisations on the distribution of discourse markers can be made as yet.

If discourse markers are not embedded they may be integrated into the matrix language approach by their attribution to a matrix structure. The remainder of the finite clause which is in the other language is then regarded as an embedded constituent. We examined proposals in this direction which are formulated in X-bar syntax and which associate discourse markers with a matrix structure called CP from the same language. In X-bar syntax such an analysis seems appealing only for clause-initial and clause-final discourse markers. In the case of clause-internal discourse markers, their linking to a syntactic structure (CP) above the finite clause (IP) level causes complications that can only be solved by assuming movement operations.

Even for the monolingual language varieties it is an intricate task, to say the least, to capture the syntactic properties of the heterogeneous class of discourse markers in terms of hierarchical structures. With the exception of subordinating conjunctions in some languages we do not find evidence of any clear impact of the discourse marking categories on the finite clause, in terms of word order or otherwise, which would argue for the existence of a higher order matrix structure such as the CP being associated with the discourse marker. In CS contexts we could not even establish a hierarchical relationship between subordinating conjunctions and clauses from another language. This argument aside, I am reluctant to regard a complicated structure like the finite clause as a frequently embedded constituent in view of the fact that constituent insertion generally tends to be rather constrained in CS.

Ultimately, the association of a language *y* discourse marker with a language *y* CP says nothing beyond the fact that the language *y* discourse marker preserves its language *y* distributional properties in relation to language *x* clauses. Now this observation also applies to discourse markers that are not indicative of a CP matrix structure. As a case in point, recall the expletive adjectives in Irish/English (*fuckin'*, *friggin'*, *bloody*). Pragmatically these function like sentence adverbs, evaluating what is being talked about. Syntactically they are attributive adjectives and as such they are embedded in Irish (ML) NP constituents. But they share with sentence adverbs the property of retaining their source language, that is prenominal, word order in CS context (see the discussion of examples (80) and (81) in Ch. 1, p. 102 and this chapter, section 2.2.3). Of course, adjectives like *fuckin'* are unlikely candidates for the Spec of CP position. Indeed, they cannot be considered as belonging to a matrix structure above and outside the finite clause level, at least not in the syntactic sense.

Therefore, as an alternative to a hierarchical matrix structure approach, and to syntactic analyses like the CP structure explicated above, we could consider the possibility that syntactic properties are directly associated with certain adverbs, conjunctions and other discourse markers in certain languages. That is, these elements do not fill a slot in a matrix structure, nor do they project a matrix structure themselves. Instead, they ‘go looking’ for their own position in or adjacent to a clause structure. So an English sentence adverb like *unfortunately*, and a Tagalog clitic particle like *ho* (respect marker) would each have their specific syntactic properties associated with them.

The idea is that syntactic information is situated in the lexicon for elements whose word order is not governed by other principles like phrase structure or government. This may seem a far-reaching conclusion but, compared with the proposal that adverbs or Topic pronouns project CP structures, it is a relatively moderate lexicalist claim. Twenty-five years ago Hasselmo proposed such a lexicalist explanation for the ‘unsubmissive’ syntactic behaviour of English adverbs in American Swedish.

Yet other rules, such as the placement of the adverb before the finite verb in the main clause (*han* finally *gick* *hem* [see (7) above]), show a certain interdependency between the choice of the lexeme and the choice of transformation rules, which, in this case, depends on the choice of an English or Swedish lexeme. (Hasselmo, 1972a: 175)¹⁵

Although it is not clear what transformation rules Hasselmo had in mind, these rules eventually lead to the placement of English sentence adverbs according to English syntax in Swedish clauses.

Codeswitching supports the idea that discourse markers are in some way independent of matrix structures. Remember the Tagalog adverbial clitics in Tagalog/English discussed in 3.2.4 above. As for monolingual Tagalog, we may conjecture that Tagalog clause structures contain a second position slot which is optionally filled by one of these adverbs. However, we cannot assume that English clauses have such a slot for these adverbs, and yet the Tagalog adverbs surface as second position clitics in Tagalog/English CS. On the other hand, the adverb’s independence from the finite clause is not total, since it has to have some ‘understanding’ of this matrix structure in order to find its placement in second position.

¹⁵ “Åter andra regel, tex placering av adverb före det finita verbet i huvudsats (*han* finally *gick* *hem*), uppvisar ett vist beroende mellan lexemval och val av transformationsregel som beroende av valet av engelskt eller svenskt lexem i detta fall” (Hasselmo, 1972a: 175).

3.4 Summary: a separate level for discourse organisation?

We have examined the relationship between the Matrix Language defined in Chapter 2 and several aspects of discourse grammar. One aspect of discourse grammar that is manifest in CS is the use of special morphemes that, using Schifffrin's (1987: 326) definition, "index an utterance to the local contexts in which utterances are produced and in which they are to be interpreted". Besides discourse markers, clause-external foregrounding strategies were discussed as another aspect of discourse grammar.

After a review of the use of discourse markers and foregrounding constructions in various CS situations, I discussed the possibility of including these in a matrix language model. In particular, the proposal by Myers-Scotton and her associates to include these aspects of discourse grammar in X-bar structures was given due attention. The Topic pronouns that precede clauses in Arabic and in CS with Arabic played an important role in the development of their CP analysis. They propose the idea that the clause-external Topic or discourse marker is part of a syntactic matrix structure called CP. This CP further contains a finite clause, and if this clause happens to be in a language other than that of the external Topic or discourse marker, it is an EL constituent within the CP matrix.

I am reluctant to adopt the idea of a supra-clausal matrix structure for the clause-external foregrounding constructions in Japanese/English and in CS with Arabic. It must be recognised that Topic-Comment is a Japanese construction just as the use of Topic pronouns is typically Arabic in the CS contexts concerned. However, it remains an intricate problem to account for these constructions and, in the case of Arabic, their syntactic properties within classical binary branching X-bar structures. Or, in other words, the proposed X-bar structures suggest more syntactic rigour and uniformity than actually exist. Besides, although we may recognise a matrix structure on this supra-clausal level, there is no independent criterion that identifies the ML. The ML can only be established if the particular clause-external foregrounding is specific to one of the languages involved. Therefore the postulation of a higher order matrix structure that includes the clause and the clause-external Topic does not yield as many useful generalisations as does the ML on the finite clause and the lower order constituent level. Another reason for reluctance concerning this analysis is that the recognition of the finite clause as a commonly embedded constituent is not in accordance with the general observation that constituent insertion becomes increasingly constrained with more complex constituents.

As for discourse markers, which are taken to comprise conjunctions, sentence adverbs, and particles, they are difficult to associate with either the matrix language or the embedded language because of their peripheral syntactic status. On the other hand, some markers have syntactic properties that link them to either their source language or the other language in a bilingual conversation.

In general there is little evidence that these function words are embedded in a frame (as is the case for argument constituents and content words). Nor is there much evidence that they project a syntactic frame (as do verbs and phrasal categories). An interpretation along the lines of the CP analysis, which links the discourse marker

to a CP structure and identifies the finite clause in the other language as an EL constituent, was rejected. Besides the objections just mentioned in connection with clause-external foregrounding, the argument was put that the CP analysis does not satisfactorily solve the problems with clause-internal discourse markers and (other) adverbs that maintain their source language word order whenever they are embedded. I made no further attempt to incorporate discourse markers in an ML model. Instead, I suggested that for some discourse markers at least, their syntactic properties are part of their lexical entries. That is, their word order is not imposed by a syntactic frame, but as independent entities they go looking for their position within, or adjacent to a sentential structure.

Van Staden's (forthc.) findings on codeswitching between Tidore and North Moluccan Malay show that various Malay elements maintain their source language syntactic properties when used in Tidore sentences. Besides the modal verbs and aspectual adverbs and conjunctions discussed in sections 3.2.5 and 3.3.4 above, this observation pertains to Malay negation particles and numerals. Van Staden notes that these codeswitching data are in accordance with one of the universals of language contact proposed in Moravcsik (1978):

A lexical item that is of the 'grammatical' type (which includes at least conjunctions and adpositions) cannot be included in the set of properties borrowed from a language unless the rule that determines its linear order with respect to its head is also so included. (Moravcsik, 1978: 113)

Van Staden's data lead her to emphasise the grammatical function as a shared aspect of the elements that defy the matrix language model. In this chapter I identified discourse grammar as a common ground for a heterogeneous set of data that challenge the concept of insertion in general as well as the Monolingual Structure Approach developed in the preceding chapter. 'Grammatical elements' as a generalisation covers the Malay modal verbs in Tidore which preserve their source language word order; conjunctions and perhaps also numerals can also be subsumed under this category. On the other hand, the generalisation 'discourse grammar' rather than 'grammatical elements' includes phenomena like extra-clausal foregrounding in the Japanese Topic-Comment structures which are not necessarily marked by specific lexical items. In addition, the generalisation 'discourse grammar' excludes many function morphemes which do not usually occur as singly embedded morphemes in clauses or sentences from another language, notably determiners and prepositions, as well as bound morphemes in general. However, as the meticulous definition of discourse markers makes clear, grammatical functions on the discourse and syntactic level overlap, with the same lexical items often functioning at both levels. Thus, an adverb like the English *finally* can function at the discourse level as a marker of textual organisation ('to conclude my talk'), as an expression of subjective modality ('he finally went home' i.e. 'I wish he'd left earlier'), or as a lexical marker of Tense ('at the end of

a series of events, he went home'). True, these uses are associated with different syntactic properties and preferred word orders but the relationship between word order and function is neither consistent nor immutable. Likewise, a causal marker such as *because* may mark factual causation ('my screen is all black because it's broken') besides epistemic substantiation ('my screen must be broken, because it's all black') and speech act justification ('do you have a spare pen? because I forgot to bring one'). Factual causation tends to be more syntactically constrained than the two other causal relationships. My claim is that those functional elements which are not from the matrix language on the finite clause level (as identified by the finite verb criterion) tend to function first and foremost on the level of discourse grammar.

At this stage we may conclude that the matrix language concept has little to offer to the analysis of extra-clausal foregrounding and discourse marking in code switching. Various modal and aspectual adverbs and (in the Tidore case) modal verbs defy the concept of an ML, which determines the occurrence and word order of all constituents in the finite clause. However, rejecting the concept of matrix language altogether on the basis of this conclusion would be throwing out the baby with the bathwater. The ML provides an indispensable anchor for the localisation of these phenomena. If the ML is identified by the same criteria as before, it becomes possible to talk about such elements as language *x* discourse markers that precede or follow language *y* finite clauses - information that may be as relevant as knowing the language of the morpheme immediately adjacent to the discourse marker. This is especially relevant for markers that have recognisable syntactic properties. After all, if the placement of an adverb is best described as 'clause-initial/final', 'clause-medial' or 'second position', then the adverb is located with respect to a syntactic structure that needs to be defined. The most efficient way to describe that structure is by referring to the grammar of one language, viz. the matrix language. Naturally, in the examples discussed above it was possible to make observations on the placement of adverbs and discourse markers only after an ML had been assumed in the first place. The matrix language approach furthermore enables us to make a cross-linguistic comparison of adverb placement in codeswitching.

The alternative, a linear approach of the type "a Tagalog adverbial clitic can come between an English subject that is *not* a pronoun and an English verb" (Sobolewski, 1982: 41) would be a roundabout way to describe the phenomenon. Although this method will eventually cover the same language facts, in this particular example it misses two important generalisations: 1) the Tagalog adverb tends to occur in second position, 2) what elements may occur adjacent to the second position adverb is determined by English grammatical rules (cf. 3.2.4 above on adverbial clitics in Tagalog/English).

Separate level for discourse organisation?

We have seen many examples where discourse markers and foregrounded constituents from one language occur adjacent to finite clauses from the other language. We may therefore also contemplate the possibility of a general discourse marking level above

the finite clause level. It would be useful if, for a particular stretch of a bilingual conversation, we could identify one language as the ML on discourse level, with the other language serving as the ML on the finite clause level. Some bilingual conversations do indeed seem to display such a pattern. On the one hand, one gets the impression that bilingual speakers continue to use discourse marking mechanisms from their in-group language (Community Language) even when they have shifted to the use of the culturally dominant language (Superimposed Language, see Chapter 11 for a discussion of these terms) for the remainder of the conversation. On the other hand, there may be a general tendency to use discourse markers from another language because these, by virtue of being CS forms, are experienced as more salient by the interlocutors (De Rooij, 1996). Now that extra-clausal foregrounding and the distribution of discourse markers do not seem to be so constrained by syntactic features of either language in a CS situation, it may be more rewarding to investigate how their use relates to sociolinguistic variables.

Unfortunately, the reality turns out to be rather complicated. In most bilingual conversations discourse markers from both languages are used. Gardner-Chloros's (1991) quantitative data on her Alsatian/French text corpus illustrate this point very clearly. As is common in contact situations, the relative status of the languages involved is unequal and consequently CS patterns are highly asymmetric. I will have more to say about this in Chapter 11 but with regard to discourse marking in bilingual conversations, consider Gardner-Chloros's Table 6.2 (1991: 164) which is reproduced below.

Note that many more French content words occur in Alsatian utterances than Alsatian content words in French utterances, whereas with respect to conjunctions and "greetings, interjections" and so on, the division between languages is strikingly more symmetric.

Breakdown of Single-Word Switches by Language and Grammatical Category

	French in an Alsatian utterance	Alsatian in a French utterance

Breakdown of Single-Word Switches by Language and Grammatical Category

Nouns	58	2
Verbs	9	2
Adjectives	6	0
Adverbs	13	1
Conjunctions	4	3
Greetings, interjections, tags, phatic markers, and other expressions	27	17

*Table 3.1. Quantitative data on Alsatian/French in Strasbourg
(Gardner-Chloros, 1991: 164)*

Since Gardner-Chloros's Strasbourg conversations include both French and Alsatian utterances, and the Alsatian and French utterances referred to in *Table 3.1* presumably contain discourse markers from both languages, there is unlikely to be a single ML for discourse grammar within a stretch of conversation. The same holds for the Moroccan Arabic/Dutch conversations to be discussed in the following chapters: a single utterance may contain discourse markers from both languages. See also Papademetre (1994) on Greek and English discourse markers in use in the Greek community in Australia. Further investigation should shed more light on this matter. As a first step we would have to make a systematic distinction between the various functions of discourse grammar, e.g. turn taking, sequencing in narratives, evaluation of the discourse or expression of attitudes toward interlocutors, in order to see whether any of these functions can be systematically related to one language in a bilingual conversation, or even in CS generally. Since it is beyond the scope of the present study to pursue this line of argument, we must be satisfied with the acknowledgement that the Monolingual Structure Approach has reached the limits of its applicability at this point.

Part II

Description of Moroccan Arabic/Dutch

Codeswitching

Bħal haduk eh *de wetenschappelijke mensen* ka-ydiru *stukjes schrijven over ontstaan van de wereld of wat dan ook, maakt niet uit*, huma ka-ydiru *eh veel moeilijke woorden gebruiken, ontzettend moeilijk te lezen*, i baš ydiru *ehm alles opschrijven eh dat is gewoon mierenneukerij*.

“For example those scientific people, they write texts about the creation of the world or whatever, doesn’t matter what. They use many difficult words, very difficult to read, just in order to write everything down. That’s just nitpicking.”

Samir, from the Nijmegen corpus.

Introduction to Part II

The following description of Moroccan Arabic/Dutch codeswitching aims to provide an account of the language data within the framework of the Monolingual Structure Approach to CS. First I will provide the relevant information about the Moroccan community in the Netherlands, the Nijmegen data corpus, and the respondents who took part in the recordings (Chapter 4). Then follows the morphological and syntactic description of the data, which will be divided in two parts. The first part concerns insertions of Dutch morphemes and constituents in MA matrices; the second part focuses on MA insertions in Dutch. The part dealing with MA/Dutch CS with MA as the matrix language comprises four chapters; Chapter 5 discusses the insertion of Dutch nouns and nominal constituents, Chapter 6 discusses the insertion of verbs and their complements. Embedded Dutch prepositional constituents and adverbs are covered in Chapter 7, and Chapter 8 deals with embedded Dutch clauses and some discourse markers. As there are far fewer instances of MA insertions in Dutch matrices, these will all be discussed in Chapter 9. In this chapter the description of MA/Dutch CS with Dutch as the ML is organised into sections corresponding to Chapters 5 to 8, with the exception that no MA verbs were inserted into Dutch clauses. A summary of the findings is presented at the end of Chapter 9, and in another format in Chapter 10.

For the reader who is not very familiar with these languages, general grammatical information on MA and Dutch is provided whenever necessary in order to appreciate the CS patterns discussed. For more basic or elaborate information on the structure of these languages I refer the reader to the following reference works: on Dutch: Donaldson (1997); on Moroccan Arabic: Harrell (1962), Caubet (1993) and, for a concise outline of MA grammar, the grammatical compendium in Otten's (1983) dictionary. In the examples cited the passages in MA are accompanied by a morpheme-by-morpheme gloss (cf. Notational Conventions on p. 1).

Like elsewhere in the present study, I indicate the respondent for each example from the Nijmegen corpus, so that the reader may link the examples to the sociolinguistic information provided in Chapter 4 (section 4). The pseudonyms are also used in order to indicate the distribution of particular CS phenomena amongst the respondents of the Nijmegen corpus.

On quantification

The grammatical description in Chapters 5 to 9 is a qualitative, rather than a quantitative analysis of the data. This choice was guided by the following considerations. Quantitative information becomes useful only if it exists for all categories that are in complementary distribution. To give an example: in order to get an impression of how common it is for Dutch content verbs to be inserted, the number of EL content verbs in itself is not very informative unless we can compare it to, say, the number of Dutch EL nouns. Then if it turns out that more nouns than verbs are embedded, we would want to know whether this difference reflects the overall frequency of content verbs as compared to nouns in the monolinguals' speech varieties. This means that in addition to counting CS instances, a word count for sufficiently large samples of the monolingual would be required. As to what would constitute a sufficiently large sample would depend, of course, on the level of detail desired.

A significant complication for quantitative analyses of the Nijmegen corpus is its heterogeneous character. The corpus includes monolingual stretches in both languages, stretches with both frequent and infrequent CS and, most importantly, there is considerable individual variation with respect to both the monolingual and CS varieties. Let us take the insertion of Dutch verbs once more as an example. On the basis of an informal impression of the data and a count of the instances of embedded verbs, I come to the conclusion that, while all respondents insert Dutch nouns, the insertion of verbs is common in the speech of some respondents, infrequent in the contributions of some others, and absent from the data of a third group of informants. Quantitative data on the corpus as a whole would therefore be pointless. The outcome would result from coincidental factors such as the composition of the respondent group and each individual's share in the recorded conversations. Given the differences between respondents, a detailed quantitative analysis would be useful only in the case of analysing the idiosyncratic varieties. Moreover, even then we would have to assume that the data from one and the same speaker are sufficiently uniform and representative of her speech behaviour. This is not so obvious as it seems, since several scholars have shown that CS patterns are influenced by characteristics of the speech event such as the topic under discussion and the interlocutors present (see Ch. 1 section 2).

Ideally we would need a comprehensive word count of the data corpus, specified for language, word class, and grammatical function, as well as for speaker, conversational setting and perhaps other sociolinguistic factors (addressee, topic). At present the Nijmegen corpus is not available in a format that would permit such a word count, if only because it is not transcribed in its entirety. Even if there were a complete transcription, it would be a tremendous undertaking to mark each word for the various classifications. The insights gained by a quantitative analysis would not justify the investment, the more so since a quantitative analysis of another MA/Dutch codeswitching corpus already exists at our disposal (Nortier 1990).

Consequently, in the ensuing discussions, remarks on the frequency of the observed phenomena will be based on impressions rather than hard statistics. However, in order to exclude non-recurrent phenomena the following minimum standards are observed: unless stated otherwise, the phenomena discussed occur at least five times in the data, distributed among at least two respondents. In the case of less frequent insertion types (less than 20 occurrences), the absolute number of tokens is given, so as to allow some comparison of different insertion types and individual respondents.

Chapter 4

The Nijmegen Data Corpus

This chapter opens with a short introduction to the Moroccan community in the Netherlands, followed by the history and general characteristics of the corpus of MA/Dutch CS data gathered at the University of Nijmegen. Afterwards I will discuss the backgrounds of the respondents who participated in the data collection in two sections, first viewing the respondents as a group, then introducing each respondent individually.

4.1 The Moroccan community in the Netherlands

According to the so-called ‘combined birth country criterion’, which includes everyone born in Morocco or having one parent born in Morocco, there were 195,536 Moroccans in the Netherlands per January 1, 1992 (Martens et al., 1994). This makes the Moroccans the second largest ethnic minority group, in terms of nationality, after the Turkish community, which numbered 240,810 individuals in 1992. The figure of 195,536 does not include unregistered illegal immigrants, obviously. Moroccan migration to the Netherlands originated in the 1960s when Dutch employers in the industrial and agricultural sectors began to recruit personnel from various Mediterranean countries, due to a shortage of unskilled labour in the Netherlands. After the economic crisis of 1973, the recruitment of foreign workers factually came to an end. However, immigration continued as Moroccan workers had their wives and children come over (family reunion) and, particularly since 1984, due to new marriages and relations (family formation) (Muus, 1993: 56). From 1975 to 1991 there was a Moroccan migration surplus of approximately five thousand per annum on average, with an increase towards the end of this period (Muus, 1993: 54). With respect to their migration history, Moroccans in the Netherlands may be divided in three broad categories: 1) an ‘older first generation’ who are now in their fifties or sixties, 2) a ‘second generation’ grown up partially or entirely in the Netherlands, some of whom have young children of their own (third generation), and 3) a ‘younger first generation’ of recent immigrants who arrived as young adults as a result of family reunion or family formation. Note that the ‘younger first generation’ also comprises the adult children of the ‘older first generation’, as far as these were raised in Morocco. In fact, quite a few immigrants from the older generation tend to have their children brought up and educated in Morocco.

Especially among the first immigrants to the Netherlands, people from the Berberophone Rif area in northern Morocco are well represented. A study of the Dutch

national statistical office shows that in 1984 sixty percent of a random sample of just over 1000 Moroccan immigrant families originated from northern Morocco, in particular from the Rif area (Centraal Bureau voor de Statistiek, 1986: 24). In addition to the speakers of Rif Berber, also known as Tarifit, there are small numbers of speakers of the other two Berber languages of Morocco, Tashelhit (High Atlas, Anti Atlas, and Sousse Valley) and, less significantly, Tamazight (Middle Atlas). It is estimated that about 70% of the Moroccans in the Netherlands speak a variety of Berber as their mother tongue (Van der Meer, 1984: 141). Both in Morocco and in the Netherlands, the speaking of a Berber language is stigmatised and associated with backwardness, although recent years have witnessed a revival of the Berber identity and ethnic self-consciousness. Since Moroccan Arabic is the lingua franca in Morocco, nearly all Berberophones have a good command of this language, with the exception of elderly women from remote areas and, in the Dutch context, Dutch born children of Berberophone parents. Apart from Berber and Moroccan Arabic, which are both restricted to oral communication, many Moroccan immigrants have a knowledge of Standard Arabic and/or French, the languages used in education and in mass media in Morocco (cf. Wagner, 1993).

The social situation of the Moroccan community is far from ideal. Economic developments led to a reduction of job opportunities for unskilled workers and a dramatic unemployment rate in migrant communities generally, and amongst Moroccans in particular (Veenman, 1994a). The Moroccan youth faces problems of poor educational achievement and delinquency, although a distinct improvement has been signalled regarding educational achievement in recent years (Veenman, 1994b; De Ruiter, 1994; Werdmölder, 1994). The ideological antagonism between Islamic and Western values is becoming increasingly accentuated. If we take all these factors into consideration, adding the inherent human tendency toward ethnocentrism, it becomes clear that Moroccans encounter unfavourable attitudes and discrimination in Dutch society. As a result, the Moroccan diaspora in the Netherlands, and elsewhere in Europe, is very much aware of its ethnic identity. In this respect it is noteworthy that the Nijmegen data corpus was collected from February 1991 to February 1992, shortly after the second Gulf war (1990-1991). This war was commonly experienced as a conflict between the Arab Islamic world and the West, and the Moroccans in the Netherlands were associated, and sometimes identified themselves, with the Iraqi dictator Saddam Hussein, despite the official Moroccan participation in the anti-Iraq alliance.¹

¹ The chapters on the Netherlands and on Morocco in *L'annuaire de l'émigration* (Bafao & Taarji, 1994) offer an introduction to various social and economical aspects of both emigration from Morocco and the Moroccan community in the Netherlands.

4.2 The Nijmegen data corpus

The present research project was initiated by Jacomine Nortier. It started in 1991 as a continuation of Nortier's earlier research in this field (cf. Nortier, 1990). As the first step she undertook the collection and ordering of a new data corpus. In 1993 I became responsible for the codeswitching project in Nijmegen, and the data that had been gathered were put at my disposal.

These data consist of audio recordings of interviews and spontaneous conversations among Moroccan immigrants and immigrants' children, in addition to reports on the immigration history and patterns of language use of each respondent. A student of Moroccan descent organised the recording sessions and also took part in most of them. He subsequently transcribed the passages that contained intra-sentential or intersentential CS. These transcripts were then checked by Jacomine Nortier. I went through all the transcripts myself, and corrected them according to my own perception where I felt it necessary.

The sociolinguistic information on the individual respondents (see below) was provided by the research assistant, if he knew the respondents personally. Additional information was obtained during the recorded conversations. In some cases, the respondents were asked to introduce themselves at the beginning of the recording session. Later, this practice proved very useful because it facilitated the identification of the speakers on the audio tapes. The information obtained was laid down in a 'sociolinguistic profile' form, an example of which is reproduced in the Appendix at the end of this book (p. 425).

A total of 23 respondents participated in the recordings, resulting in approximately twelve hours of taped conversations. The respondents were aware of their being recorded, and they were given a modest financial reward for their cooperation. Not all recordings were equally useful. I retained only the clearest recordings and the respondents who contributed substantially to the CS data. Sometimes the volume was too low. One respondent spoke very little, and almost exclusively in MA. Two conversations were skipped because it was not possible to make a reliable transcription. In both of these, five men in their early twenties participated, and in the heat of the discussion it often occurred that more than one spoke at the same time. At various points it was too difficult to identify the respondents post hoc by their voices. Eventually, nine male and six female respondents remained, distributed over 10 conversations (approximately nine hours).

The data are heterogeneous in many ways. This variation is apparent in the sociolinguistic backgrounds of the respondents, their speech behaviour in terms of language choice, and the conversational settings. Detailed information on both the conversations and the individual respondents is given below.

4.3 The respondents as a group

The following information concerns the respondents as a group, comparing them with respect to their migration history and linguistic background. All the respondents' names are pseudonyms.

Age and migration history

Most of the respondents were adolescents ranging from 17 to 29 years of age, with the exception of three individuals; two younger and one older, cf. *Table 4.1*. Samir's younger siblings Nawal and Abdelkrim were 11 and 13, respectively, when they were recorded. The older respondent is Maryam's mother Hayat, who was 39.

Table 4.1. Respondents by age (years:months)

pseudonym	age at recording	age at immigr.	duration of stay
Nawal	11:4	local born	11:4
Abdelkrim	13:1	local born	13:1
Younes	18	3:4	14:8
Samir	19-20	6	13-14
Jamal	20	7	13
Fatima	21	16	4: ?
Mimoun	21	20*	2+1
Maryam	21	2:6	18:6
Mustafa	25	20	4:6
Najib	24	21	3:4
Abdellah	25	10	15
Hocine	26	21	5
Warda	26	24	2:?
Zineb	29	26	3
Hayat	39	20	19

Apart from Hayat, the respondents are divided between the ‘second generation’ and ‘younger first generation’ categories. The division between these two categories is not very distinct, however, since the ‘second generation’ includes individuals who are not Dutch born but who were brought to the Netherlands at an early age, while the ‘younger first generation’ includes immigrants’ children who joined their parents when they were young adults. Only Nawal and Abdelkrim were born in the Netherlands; Maryam, Samir, and Younes arrived at pre-school age, Jamal at the age of 7, and Abdellah at the age of 10. For Mimoun the situation is more complicated, as he moved three times between Morocco and the Netherlands (see *Mimoun* in section 4 for the details below). Fatima was 16 when she came to the Netherlands; all the remaining respondents were 18 years or older when they arrived. Depending on the respondents’ age at the time of the recordings and their age of immigration, the amount of time spent in the Netherlands varies considerably, between less than three years in the case of Warda, and 19 years in the case of Hayat.

Linguistic background

a) language skills

The respondents’ competence in Moroccan Arabic and Dutch is obviously related to the amount of time spent in each country and to their age upon arrival in the Netherlands. Generally speaking, the younger a person was when he or she immigrated, the better (s)he acquired the new language. With the exception of Mustafa, all selected respondents were capable of sustaining a conversation in either language, though with varying degrees of fluency. Language abilities have been appraised only by means of self-assessment. The respondents were asked whether they were more confident in speaking either language. Jamal and Abdellah, who arrived in the Netherlands at the age of seven and ten respectively, are closest to being ‘balanced’ bilinguals, that is to say, having no clear preference for either language. The respondents were also asked which language they preferred to speak. None preferred a different language than the language (s)he spoke best. Those had who immigrated at an earlier age, or where born in the Netherlands, had a preference for Dutch; all respondents who had arrived at the age of 16 or older were more confident in speaking MA.

Table 4.2. Respondents by language preference

pseudonym	age at immigr.	preferred language
Nawal	Dutch born	Dutch
Abdelkrim	Dutch born	Dutch
Maryam	2:6	Dutch
Younes	3:4	Dutch
Samir	6	Dutch
Jamal	7	Dutch/MA (Tashelhit)
Abdellah	10	Dutch/MA (Tashelhit)
Fatima	16	MA
Mimoun	20*	MA (Tashelhit)
Hayat	20	MA
Najib	21	MA (Tarifit)
Hocine	21	MA (Tarifit)
Warda	24	MA
Zineb	26	MA
Mustafa	20	MA

Note that ‘language preference’ is not meant to imply that the respondents did not like to speak the other language; in fact, some indicated that they wanted to develop their skills in their weaker language. Nor does language preference imply perfect mastery of the preferred language.

b) Berber

Although Berber is not considered in this particular study, we note that five of the informants speak a variety of Berber as their first language: Najib and Hocine grew up in the vicinity of Al Hoceima and Nador, two largely Berberophone (Tarifit) cities on the Mediterranean coast. Abdellah, Jamal and Mimoun are of Soussi descent, and speak Tashelhit besides MA. The MA of the Tashelhit speakers does not differ

noticeably from that of the respondents for whom MA is their first language. The Tarifit speakers Najib and Hocine speak MA somewhat less fluently than the other respondents who grew up in Morocco. Their MA is characterised by a slower speaking rate, more hesitation phenomena and, interestingly, more lexical influence of the written language Standard Arabic.² Mustafa, who is an Arabophone from the Rif region, reports that he has acquired Tarifit during his last years in the Netherlands, so as to be able to communicate with fellow Moroccans.

In view of the Moroccan sociolinguistic context it is not always easy to assess a person's mother tongue. One factor is that, because of the minorisation and stigmatization of Berber languages, some individuals are reluctant to admit that they speak a Berber language, or used to speak it in their childhood days. A more significant complicating factor is the internal migration in Morocco, particularly from Berberophone rural areas to Arabophone urban centres, as a result of which many children grow up bilingually. Many families are made up of Arabophones who do not speak Berber, as well as Berberophones who speak both Berber and MA. The Hamadi family (see below) is a telling example of this. No Berber was spoken during the data recording, which indicates that these Berberophone respondents are fully capable of keeping MA and Berber apart as distinct language systems.

c) Arabic dialects

The Arabophone and Berberophone respondents originate from various parts of Morocco, accordingly displaying a range of dialectal varieties of Arabic. Traditionally, Moroccan Arabic dialects are divided in three main groups: the so-called Urban (Mdini) and Mountain (Jebli) dialects, both of which date back to the earliest Arab settlements in the 7th century; and the so-called Rural or Bedouin dialects that result from the invasion of the Beni Hilal and Ma`qil tribes in the 12th and 13th centuries (cf. Colin, 1945). The main division is between the first two groups, also known as 'pre-Hilali', and the latter ('Hilali'). Fez, Rabat, Salé, Tétouan, Taza and El-Ksar are the main centres of the Urban dialect group; Mountain dialects are found in the northwestern part of Morocco, from the Strait of Gibraltar to Ouezzane, and from Ouezzane to Taza. Bedouin dialect speakers populate the plains.

² In view of the various nationalist opinions, the Arabic language skills of Berberophones are a sensitive and controversial issue. I have no indication that the Berberophones' achievement in Standard Arabic does not measure up to that of Arabophones (cf. Wagner, 1993: 175-6). With respect to MA, however, I have the impression that Tarifit speakers in general can be distinguished from native speakers. Circumstantial evidence comes from my observation that their MA is often easier to understand for me as a non-native speaker. In addition to those mentioned above, I noticed the following distinctive traits: infrequent (unproductive?) use of the medio-passive prefix *tt-/t-* in verbs; wider distribution of the analytic, as opposed to the synthetic, genitive construction (Boumans, 1994); infrequent use of idiomatic expressions; errors in gender and number agreement. But it can be argued that the two former traits fall within the scope of dialectal variation in MA.

Three Bedouin dialect groups have been distinguished. According to Laghaout's (1995) nomenclature these are 1) Aroubi along the Atlantic coast from Asila to Essaouira; 2) Bedoui in the East in the basin of the Moulouya river and on the plateaus east of the river, and 3) Hassaniya in the Saharan region. Note that the 'Bedouin' dialect regions include many urban centres including Casablanca, by far the country's largest city. It is questionable, however, to what extent the traditional classification and geographical distribution of Arabic (and Berber) dialects still apply. In the past 50 years, urbanisation, increased mobility and the influence of radio and television have stimulated the emergence of a Moroccan Arabic Koine based on the Bedouin dialects of the Atlantic coast (Laghaout, 1995).

On the basis of their place of origin in Morocco or, in the case of those who grew up in the Netherlands, their parents' place of origin, the respondents of the CS corpus can be divided over three dialect groups: 1) Mountain/Urban, 2) Aroubi, or the Atlantic Coast Koine, and 3) Bedoui or 'East Moroccan'. *Table 4.3* shows the respondents' MA dialect and, where appropriate, their Berber mother tongue.

Table 4.3 orders the respondents according to the place or region of origin, roughly describing a line from South to Northwest, and from Northwest to Northeast. The two Tarifit speakers interrupt this line. I gain the impression that their MA is more similar to the Atlantic Coast Koine than to the neighbouring Mountain or Bedoui dialects. For the Hamadi siblings, the two youngest of whom were born in the Netherlands, Oujda/Algeria is given as place of origin because they learnt Arabic primarily from their Algerian born mother (see below for details). However, the Algerian Bedouin dialect most likely spoken by the mother is probably a continuation of the East Moroccan group (Ph. Marçais, 1957: 236).

Table 4.3. Respondents by MA dialect group

pseudonym		MA dialect group(Berber)
Abdellah	Sousse ³	(Tashelhit)
Jamal	Sousse	(Tashelhit)
Mimoun	Ouarzazate	(Tashelhit)
Zineb	Casablanca	Atlantic Coast Koine
Hayat	Fez	
Maryam	Fez	
Fatima	Larache	
Hocine	Nador	
Najib	Al Hoceima	
		(Tarifit)
		(Tarifit)
Warda	Tétouan	Urban and Mountain
Mustafa	Zerqat (Rif)	
Samir	Oujda/Algeria	East Moroccan or Bedoui
Younes	Oujda/Algeria	
Abdelkrim	Oujda/Algeria	
Nawal	Oujda/Algeria	

One feature that Warda's and Mustafa's varieties of MA have in common which distinguishes them from the Atlantic Coast Koine, is the pronunciation of the preposition *l* (direction) as [n] before NPs (not before pronominal suffixes). Warda's speech has a distinctively Urban flavour due to the unvoiced pronunciation of the phoneme /q/ in many words, notably [qal ~ yqul] "to say". Mustafa and all the other respondents have the voiced pronunciation [gal ~ ygul] associated with Bedouin dialects.

The East Moroccan bedouin dialect of the Hamadi siblings differs from the western varieties through the use of the third person masculine singular object suffix *-eh* in

³ For Abdellah, Jamal and Mimoun, their written sociolinguistic profile mentions Agadir as their place of origin, but I suspect that this is meant as an approximate indication for the Sousse region, or even South Morocco.

place of *-u*, gender distinction in the second person singular of perfect verbs and the absence of the aspectual prefix *ka-* with imperfect verbs. In the recorded material the Hamadi brothers actually oscillate between East Moroccan and Atlantic Coast Koine forms.

Again, I would like to stress that dialect boundaries are blurred by cross-dialectal influence and koineisation processes, both in Morocco and, probably, in the Netherlands.⁴ This does not necessarily make dialect classification easier, as it results in a lot of variation, at least in the short term. Since no study was made of the dialectal features in the respondents' speech, all remarks on this topic are based on impressions only. Nevertheless, we need to be alert to (possible) dialectal variation since it complicates the investigation in the Dutch context. In many cases it will be tempting to ascribe remarkable observations and idiosyncrasies of MA spoken in the Netherlands to the influence of MA/Dutch bilingualism, but we cannot exclude the possibility that these phenomena could be part of MA dialectal variation.

Dutch dialects are considered of less importance, since most of MA/Dutch CS concerns the insertion of Dutch elements in MA structures, so that chiefly MA syntactic and morphological structures are discussed, as we will see below. Regional or social variation in Dutch is only relevant for the respondents who grew up in the Netherlands. The Hamadi siblings from the southern village of Goirle (south of Tilburg) speak Dutch with a recognisable Brabant accent. The other respondents' speech is not readily associated with a dialect region in the Netherlands, although regionalisms may be discovered upon closer investigation. Regarding those who immigrated as adolescents or adults, on the other hand, non-standard forms and irregularities that are associated with second language acquisition should be taken into account.

d) other languages

Many of the respondents are highly educated and studied foreign languages at school. Those who went to school in Morocco have a reasonable to good command of the country's literary languages French and Standard Arabic, depending on the amount of schooling. Moroccan intellectuals are often oriented toward either one of the literary languages due to personal preferences, social milieu or, more significantly, professional vocation. (More technical subjects such as medicine or engineering are only taught in French.) French is also used in informal conversations among Moroccan intellectuals, often in the form of MA/French CS; see, for instance, Abbassi (1977), Slaoui (1986), and various articles by Bentahila & Davies (1983, 1991). MA/French CS also occurs in the Nijmegen corpus, in the speech of respondents who were highly educated in Morocco (especially Mimoun and Zineb). Standard Arabic, on the other hand, is not often used as a spoken language in informal contexts.

⁴ The situation in the Netherlands is different because the Tarifit Berber speakers constitute the majority of the Moroccan community and because eastern Bedoui speakers form a large part of the Arabophones.

‘Diglossic’ switching between Moroccan Arabic and Standard Arabic primarily concerns the use of learned vocabulary and conjunctions from the literary language. In these styles the verbal system tends to remain recognisably MA (cf. Heath, 1989; Youssi, 1992), except in very formal settings; see, for instance the use of Standard Arabic in radio broadcasts (Forkel, 1980). The use of Standard Arabic vocabulary is represented in the data corpus by the speech of Najib, Hocine and Warda. Some of the respondents who were educated in Morocco had also studied (some) English at school. Only Warda, the woman from Tétouan, reports to know Spanish; Fatima reports knowledge of Egyptian Arabic. Egyptian Arabic is accorded higher prestige than MA, and particularly Moroccan women learn it from television and popular songs.

The respondents who went to school in the Netherlands learnt English, sometimes in addition to German or French. As yet, the role of English in the Netherlands is not as important as Standard Arabic or French in Morocco. Conventionalized, often slangy English expressions (*fuck off!* etc.) and technical terms in specialized jargons (jazz musicians, linguists, drug users, etc.) are recurrent features of contemporary Dutch, while English terms and phrases are much used to flavour the speech of adolescents. Still, English is not used for longer stretches of informal discourse in mainstream Dutch society.

It should also be noted that some of the respondents were recruited from amongst the students of the Department of Languages and Cultures of the Middle East at the University of Nijmegen, where they studied Standard Arabic as a major or minor subject. This is a factor of influence on the MA, especially of those who grew up in the Netherlands (Samir and Maryam). They tend to supplement lacunas in their knowledge of MA with the Standard Arabic they learnt at university. In this respect, their MA is similar to that of the Tarifit speakers Hocine and Najib.

4.4 The individual respondents

The information on the respondents that was gathered by Jacomine Nortier and her assistant is presented here for each of the 15 selected respondents individually. Most of the information was provided by the respondents themselves and was not checked for accuracy. This should be kept in mind particularly when considering the information on language abilities, since some individuals are more inclined to modesty than others.

Four of the respondents, to whom I refer as the Hamadi family, are siblings. In order to avoid redundant information I will present their common immigration history and linguistic background first. Except for the Hamadi siblings, the respondents are introduced in alphabetical order; their age is indicated between brackets, as in *Abdellah* (25).

The Hamadi siblings

Samir, Younes, Abdelkrim, and Nawal were recorded together in one conversation while their mother Fatiha was also present (her monolingual contributions are not studied). Samir, the eldest, also participated in all but one of the other recording sessions.

The father of the Hamadi family is a Rifian Berber and speaks Tarifit. The mother is also of Rifian descent, but she was born and raised in western Algeria near the town of Sig, where her parents had emigrated. She did not learn Tarifit until after her marriage in 1966, when she came to Morocco to live with her husband in the Berberophone village of Beni Said. Samir was born there in 1971. Soon after that the family moved to the Arabic speaking town of Oujda, close to the Algerian border, where Younes was born. In 1977 the family moved to the Netherlands, after which Abdelkrim and the only girl, Nawal, were born.

None of the children spoke Tarifit; only Samir reported to have passive skills in this language, and remembered that it was spoken at home in his childhood days. Their Arabic has a distinctively ‘eastern’ accent, from the Moroccan point of view. All spoke Dutch better than MA and also preferred speaking Dutch. They spoke MA with their parents and sometimes amongst themselves, the younger ones tending to use more Dutch. They also used MA in contacts with Moroccan friends. Although the younger siblings grew up entirely or almost entirely in the Netherlands, and presumably spoke less MA with friends, they did not necessarily speak it less fluently than Samir. The younger siblings still lived with their parents with whom they used MA in daily interaction. My impression is that Nawal, Abdelkrim and Younes show less hesitation and self-corrections, although their MA vocabulary is restricted. Also, the younger siblings’ speech seems to be closer to the dialect of their mother.

Samir Hamadi (19-20)

Arriving in the Netherlands when he was six, he attended primary school and various types of secondary schools until he ended up at the University of Nijmegen. At the time of the data gathering he was in his third year of Arabic Language and Literature. As a student, he lived on his own. His friends had Dutch, Moroccan and other national backgrounds. Dutch was the language he spoke best, followed by MA. He learnt English, French and German at school and Standard Arabic at university; he further reported a passive knowledge of Tarifit.

He watched Dutch TV programmes, read Dutch newspapers and books and spoke Dutch on most occasions. He spoke MA with his parents and with some Moroccan friends. With his brothers and his sister he spoke both languages, but mostly Dutch with the two younger ones.

His MA shows signs of imperfect acquisition, in particular hesitation phenomena and frequent repairs. He supplements his restricted MA vocabulary with Dutch and the Standard Arabic learnt at university. Syntactic and morphological variation seems to indicate that his variety of MA was not stabilised (inconsistent use of the aspect marker *ka-*, which is an influence of the MA Koine on the East Moroccan variety

spoken at home; anomalous variation of quantitative expressions, cf. Boumans, 1994). At the time of the data collection, Samir was very much concerned with his Moroccan background and his identity in Dutch society. This also involved an increased interest in MA, which he was eager to practise in order to improve his skills. He also had the desire to learn Berber. Regarding his competence in Dutch, Samir is not readily distinguishable from native speakers of comparable educational background, although detailed investigation of his Dutch would probably reveal some differences.

Younes Hamadi (18)

He came to the Netherlands as a pre-schooler. He attended primary and secondary school (11 years in total), and left the latter without a diploma. At the time of data recording he was following part-time education. He lived with his parents and younger siblings. His friends were of Dutch and Moroccan origin. At school he had learnt some English, French and German. Dutch was the language he used in nearly all situations, but he spoke MA with his parents and with his older brother Samir; with Moroccan friends he spoke either Dutch or a CS variety of MA and Dutch.

Abdelkrim Hamadi (13)

13 years old, born in the Netherlands. During the recording period he was in the second year of secondary education. He lived with his family and had friends from Dutch, Moroccan and Turkish ethnic backgrounds. At school he learnt English, French, and German. In the weekends he went to the mosque for Islamic instruction, which included the teaching of Standard (Classical) Arabic. He spoke MA with his parents and sometimes with his older brothers; with Moroccan friends he spoke either Dutch or a CS variety of MA and Dutch. In all other situations he used Dutch.

Nawal Hamadi (11)

Nawal was in the last year of primary school. She lived with her family and had friends of Dutch and Moroccan origin. She learnt some English at primary school. She spoke MA with her parents and sometimes with her older brothers. She reported speaking MA or a mixture of MA and Dutch with friends in about 25% of the occasions.

Abdellah (25)

Abdellah had spent his first ten years in Morocco, in or near Agadir, in the region of Sousse (see n. 3 on p. 165 above). He had had five years of (primary) education in Morocco, six years (1 primary, 5 secondary) in the Netherlands, and he was then in the third year of his training as a laboratory analyst. He lived on his own in Amsterdam. His friends had various ethnic backgrounds, although most of them were Dutch. He had learnt French at school while he was in Morocco; in the Netherlands he learnt German and English. He said he spoke Dutch best, followed by MA and Tashelhit Berber, with English and German sharing third position. He spoke MA and Tashelhit with his parents, brothers and sisters and with Moroccan friends. With other

friends and colleagues he spoke Dutch, French, or English. He watched Dutch TV, read Dutch periodicals and performed most of his language tasks in Dutch, although he reported counting sometimes in MA.

Fatima (21)

A native of Larache, Fatima was 16 when she came to the Netherlands, about four and a half years prior to the data collection. She had left primary school in Morocco prematurely after five years. In the Netherlands she attended a domestic science school for three years, followed by one year's training in clothing manufacture. For six months she had been following Dutch language courses at the University of Nijmegen in order to improve her language skills. She lived with her parents. She had mainly Moroccan women friends and a few Turkish ones, but no intimate contacts amongst the Dutch. Asked which languages she spoke best, she rated 1) MA, 2) Dutch, 3) Standard Arabic and 4) Egyptian Arabic. She spoke MA with her relatives and friends. She did not watch much TV and read very seldom, although she did read romance novelettes both in Arabic and in Dutch. Characteristic of her CS behaviour is the frequent insertion of Dutch modal adverbs.

Hayat (39)

Born in Fez in 1953, she was the oldest respondent. In 1973 she joined her husband in the Netherlands, together with her then two and a half-year-old daughter Maryam. She had not pursued any formal education in Morocco. In the Netherlands she had followed Dutch language and sewing courses and she had also obtained a hairdresser's diploma, although she never actually worked as a hairdresser. Her friends were of both Moroccan and Dutch origin, most of them being Dutch. She spoke both MA and Dutch with her two children,⁵ whereas with her husband and relatives she spoke mainly MA. She watched Dutch TV. According to her own judgment, her Dutch was poor, but she enjoyed speaking it.

Hayat's case is quite different from the general image of the older first generation Moroccan women, as she spoke Dutch rather fluently and seemed to have more contacts with Dutch society. Apart from personal characteristics, this is probably due to the fact that she followed her husband at a much younger age than the average first generation woman. Also, for years after their immigration Hayat and her family used to live in a small town where they were the only Moroccans, so they had many contacts with Dutch neighbours, children's playmates et cetera. (See also *Maryam* below.)

⁵ In one conversation in which she talked with Maryam many of Hayat's utterances, including longer stretches, were entirely in Dutch. Maryam spoke mainly Dutch. It is not clear, of course, to what extent this recording represents the speech habits of mother and daughter.

Hocine (26)

Hocine is a Rifian from a small town near Nador. He had been in the Netherlands for five years. After his secondary school (baccalaureate) in Morocco, he started studying Arabic at the University of Nijmegen. He had almost completed this study at the time of data collection, and he intended to work as a translator. In Morocco he learnt Standard Arabic, French and English at school. Asked which language he spoke best, he said he spoke Tarifit Berber, (Standard) Arabic and Dutch well, and English and French reasonably. He did not mention MA, but he spoke it fluently in the conversations in which he took part, although his MA contains much Standard Arabic vocabulary. He lived on his own and had many contacts with people of various nationalities. With his parents and siblings he spoke Tarifit; with other relatives he spoke either Tarifit or MA. With friends he spoke Dutch besides Tarifit and MA. He watched current affairs and sports programmes on Dutch and French TV and wrote and read in Dutch as well as in Standard Arabic. His MA/Dutch CS contains several insertion types that are very rare or absent in the speech of the other respondents.

Jamal (20)

Born in the Sousse region in southern Morocco (see n.3 on p. 165), he was seven when he came to the Netherlands. He had finished school and was following a vocational training in the sphere of social welfare. He shared an apartment with a Moroccan friend. The majority of his friends had a Moroccan background; the others were Dutch. In Morocco he had learnt French, and in the Netherlands he learnt English at school. In addition he wanted to learn Standard Arabic. He reported being able to speak Dutch and MA equally well, and these were also the languages he preferred to speak; Tashelhit occupied second position. He spoke Tashelhit and MA with his parents and with his brothers and sisters; with the latter he sometimes spoke Dutch. With some friends and colleagues he spoke MA or a mixture of MA and Dutch; with other friends he spoke only Dutch. He watched Dutch TV programmes and read Dutch newspapers. He read Arabic books with the help of a dictionary.

Maryam (21)

Maryam, Hayat's daughter, was born in Fez and came to the Netherlands at the age of two and a half. As a student of Arabic at the University of Nijmegen, she was repeating the first year's programme. She lived with her parents. Her friends were Dutch as well as 'Dutch of Arabic origin'. Since she started studying Arabic, she had become acquainted with 'Arabs'. She learnt MA from her parents, and reported that she spoke it reasonably well. At school she learnt foreign languages. She spoke Dutch most of the time with her parents, and she spoke exclusively Dutch with her younger brother. With friends and in nearly all other situations she spoke Dutch, but she remarked: "I speak Moroccan when I feel Moroccan, when I like to shock or intimidate or when it is necessary". (See also Hayat.)

Mimoun (21)

Mimoun is from Ouarzazate, southeast of Marrakesh. From his seventh to his ninth year he had lived in the Netherlands where he attended primary school. Then he was sent back to Ouarzazate where he attended another two years of primary education. Proving to be a very successful student, he was selected to attend elite boarding schools for secondary education in Zagora and in El Jadida with a grant from the Moroccan government. At the time of the data gathering, he had been back in the Netherlands for only 19 months. He was pursuing two studies at two universities at the same time, viz. chemistry and chemical technology. For five months he had rented a room on his own. He had friends of various ethnic backgrounds. The languages he said he spoke best were MA and French, followed by Standard Arabic and English and Dutch sharing third position. He mentioned Tashelhit as one of the languages he preferred to speak, along with MA and French. With his parents and relatives he spoke Tashelhit and MA; with friends he spoke MA, Dutch, English and French. He watched mainly English and French TV channels, read books and periodicals in these languages, but also reported reading many Dutch newspapers and magazines. He used to write poetry and song texts in English. His speech is characterised by frequent MA/French CS (which is not investigated in the following description).

Mustafa (25)

Mustafa is an Arabophone from a small place called Zerkat, southwest of Al Hoceima in the western Rif area. He had been in the Netherlands for four and a half years, and did not speak much Dutch at the time. He had not received any formal education in Morocco, where he had worked as a chicken slaughterer. After three years of work in the Netherlands he was dismissed because of his poor mastery of Dutch, he reported. He became aware of his disadvantaged position on the job market and decided to take up schooling. He had just started a Dutch language course and he was following a technical training course where he learnt to repair sewing machines and the like.

He spoke very little Dutch during the recorded conversation and, accordingly, he did not provide much CS data. His part in the data set is retained because he contributed some interesting CS examples that may be typical of beginning learners of Dutch as a second language. Moreover, his data serve as a sample of monolingual MA. The conversation with Mustafa is particularly apt for this purpose because of the length of the recording, and because he speaks slowly and clearly, and in a repetitive style.

Najib (24)

Najib is a Berberophone from the region of Al Hoceima. He had been living in the Netherlands for three years and four months. He obtained his secondary school diploma (*baccalauréat*) in Morocco and studied Middle Eastern Studies at the University of Nijmegen. He lived on his own. The majority of his friends were Moroccans, most of them Berberophones, but he also maintained contacts with Dutch. At school he learnt French, Standard Arabic, and English. He said he spoke Tarifit

Berber best, and Standard Arabic second best, with MA in third position followed by Dutch and French. Remarkably, Standard Arabic did not figure among the languages he preferred to speak (1. Tarifit & MA, 2. Dutch). He spoke Tarifit with his parents and siblings; with friends he spoke Tarifit, MA and Dutch. He watched ‘intellectual’ TV programmes on Dutch, Arabic, and French channels. He reported reading Arabic texts more often than Dutch.

Warda (26)

Warda had been living in the Netherlands for less than three years. She finished school in Tétouan, and studied biology at a Moroccan university for two years. At the time she was studying Arabic at the University of Nijmegen ‘with the purpose of learning Dutch’, as she stated. She lived with her parents, and most of her friends were Moroccans. She learnt Standard Arabic, French, English and Spanish while she was in Morocco. Standard Arabic and MA were the languages she reported to speak best; then French, followed by Spanish, Dutch and finally English. She used Arabic or MA in most situations and sometimes French. She spoke MA with her parents. With her siblings she spoke MA and occasionally a variety of MA/French CS. She read novels in French, Spanish, Arabic and, less frequently, in Dutch although she did read Dutch newspapers. She did not watch much TV.

Zineb (29)

Zineb is from Casablanca. She gained a Moroccan university degree (*licence*, approx. B.A.) in French language and literature before coming to the Netherlands, three years prior to the data gathering. In the Netherlands she had followed a course in ‘computer science and business administration’ and she had resumed her study of French at the University of Nijmegen. She wanted to become a French teacher. She lived on her own in Nijmegen, maintaining many contacts with Moroccan and Dutch friends. In Morocco she learnt French, Standard Arabic, English and Hebrew. She spoke French and MA better than Dutch. With relatives she always spoke MA; with friends either MA or Dutch. At the French department she usually spoke French with the lecturers. As a student of French literature, Zineb read many French books, although she read Dutch newspapers. She also watched Dutch TV, in addition to TV5, a French channel. Apart from MA/Dutch CS, her MA speech contains many insertions from French.

In the following chapters I will focus on the grammatical aspects of their MA/ Dutch CS. Naturally, the amount of Dutch used is proportional to the amount of time the respondent spent in the Netherlands, and the Dutch elements that are embedded in MA structures are both more frequent and more varied qualitatively in the speech of those who are the most fluent in Dutch. However, it will become apparent that, despite the respondents’ heterogeneous language backgrounds, their varieties of MA/ Dutch CS have much in common too. On the other hand, characteristics of CS behaviour that turn out to be restricted to particular speakers cannot always be related to their language background or migration history. Sometimes the same idiosyncratic

feature is shared by a number of respondents with very diverse backgrounds, as in the case of verb insertion (Chapter 6). Generally speaking, the Nijmegen corpus is less suitable for investigating the correlations between sociolinguistic factors and CS patterns. Since the 15 respondents differ from each other in so many respects (age of immigration, duration of stay, education, and mother tongue) there is a lack of internally homogeneous groups that can be contrasted with each other with respect to any particular feature.

Chapter 5

Dutch Nouns, Nominal Constituents and Adjectives in MA

This chapter deals with the Dutch nouns, nominal constituents and adjectives used by respondents in their MA clauses, or in other matrix structures such as nominal or prepositional constituents. The larger part of the chapter will be concerned with embedded nouns, the most frequently occurring category of embedded material; section 1 examines many aspects of embedded nouns and how they function in their new MA environment. The investigation focuses on the question whether embedded nouns really function in the same way as nouns from the matrix language, as the Monolingual Structure Approach predicts. Then I will briefly address the insertion of Dutch nouns that are modified by a Dutch attributive adjective (section 2), and proceed to adjectives that are embedded independently from Dutch nouns (section 3). The discussion will be briefer concerning adjectives than nouns since the former are few in number and largely restricted to the syntactic position of predicate. Embedded nominal constituents (NPs) are the final insertion type treated in this chapter and will be dealt with in section 4. The crucial question here is whether the kinds of embedded constituents sufficiently resemble NPs in Dutch for NP insertion to be seen as a recurrent feature of MA/Dutch CS. Section 5 presents an overview of the syntactic functions in which embedded Dutch nouns and NPs occur in the examples cited in this chapter and section 6 discusses the tendency to omit certain prepositions before Dutch nouns. A summary of the findings concludes this chapter.

5.1 Insertion of nouns

The examination of noun insertion is ordered as follows: the grammatical categories of gender, number, and definiteness are discussed in the sections 1.1 to 1.5. Then the modification of embedded nouns by means of quantifiers, possessives, adjunct prepositional phrases, complements, attributive adjectives and relative clauses is addressed in the sections 1.6 to 1.11. Section 1.12 deals with interrogative forms that occur mainly in Samir's contributions to the data. This ordering is inevitably arbitrary to some extent: number is an aspect of quantification, while the notions of definiteness, quantification, and modification are interrelated in many ways (cf. Caubet, 1993 vol. II on determination and quantification in MA).

5.1.1 Grammatical gender

Both MA and Dutch have a two-gender system. In MA, the grammatical genders are called feminine and masculine and coincide with natural gender for animate nouns. Apart from a small number of exceptions, feminine nouns are marked by the suffix *-a*. In addition, nouns that end in /t/ are sometimes associated with feminine gender, although this is subject to dialectal variation. MA verbs and adjectives and anaphoric pronouns show gender agreement in the singular. Some adjectives sometimes display gender agreement in the plural when referring to humans. The remote demonstrative pronoun has the paradigm {*dak*, *dik*, *duk*} for masculine singular, feminine singular and plural respectively, but in many varieties there is a tendency to generalize either *dik* or *dak* for both genders and numbers, at least for inanimates (cf. Caubet, 1998).

In Dutch, historical feminine and masculine gender are merged into one ‘common’ gender, which is opposed to neuter. Gender is not morphologically marked on the noun, except for derivational suffixes, which are associated with either common or neuter gender. Gender is primarily marked on the definite article, which is *de* for common and *het* for neuter in the singular, and *de* for all plurals.

Concerning the examination of gender assignment, only verbal and adjectival and agreement pronominal were considered. The demonstratives do not constitute a reliable criterion throughout the corpus because of the dialectal variation, as noted above. Most inserted nouns designate inanimate objects or concepts. These seldom assume the thematic role of Agent and hence, as a result of the correlation between subjecthood and agentivity, they occur less often as Subjects. Consequently, relatively few contexts could be investigated for gender assignment. The corpus as a whole shows a tendency for inanimate Dutch nouns to trigger feminine gender agreement in MA, but it is possible that this reflects an idiosyncratic feature of some speakers. The examples (1)-(3) show feminine agreement for Dutch nouns, where only in the case of *agenda* in (3) can the assignment of feminine gender be associated with the noun’s phonological shape, viz. the final *-a*. (4) and (5) are examples of masculine concord, which appears to be a minor pattern. A subscript _i indicates the MA gender agreement in verbal forms and pronouns.

feminine agreement

- (1) ma ḥna sakn-in f *dorp*_i, hiya_i ṣḡir-a_i, fi-ha_i ḡir l-huḷanḍi-yin
 while 1PL live-PART-PL in village 3F small-F in-3F only DEF-Dutch-PL
 “Now that we live in a village, which is small, and only Dutch live there.”
 (Warda)

- (2) *kayn-a_i¹ wahed video-cursus_i binnenkort in april*
 EXIST-F INDEF video-course soon in April
 “There’s a video course soon, in April.” (Jamal)
- (3) *hetta nta, agenda_i ſend-ek ſamr-a_i dima ya ſahb-i*
 also 2M diary at-2 full-F always VOCATIVE friend-1SG
 “You too, your diary’s always filled, my friend.” (Jamal)

masculine agreement

- (4) *ra-h kamer_i, ka-y-sekn-u fi-h_i tlata žuž, minimum*
 PRES-3M room ASP-3-live-PL in-3M three two minimum(French/Dutch)
 “It’s a room in which at least two, three people live.” (Najib)
- (5) *feyna l-bewijs_i, feyna huwa?*
 where DEF-evidence where 3M
 “Where’s the evidence, where is it?” (Zineb)

MA feminine and masculine gender assignment is independent of grammatical gender in Dutch: the nouns *dorp* in (1) and *bewijs* in (5) are neuter, and the Dutch nouns in (2), (3) and (4) have common gender. Not surprisingly, natural gender is respected for Dutch nouns in MA.

- (6) *kifaš vrouw_i ma ſend-ha_i-š ziel u man wel?*
 how woman NEG at-3F-NEG soul and man AFFIRM
 “How [is this possible] a woman doesn’t have a soul, and a man does?”
 (Samir)
- (7) *u y_i-gul-l-ek oom_i dyal-ek xesş-ek l-bent dyal-ek ma*
 and 3-say-to-2 uncle of-2 must-2 DEF-daughter of-2 NEG
t-xerrež-ha-š
 2-make·go·out-3F-NEG
 “And (suppose) your uncle says to you: you shouldn’t let your daughter go out.” (Samir)

¹ The MA existential marker *kayen* has the feminine and plural forms *kayna* and *kaynin*. It does not consistently show agreement, though, and in many speakers’ varieties *kayen* is invariable. Therefore, *kayen* plus a Dutch noun was not considered an instance of masculine gender assignment. This is the only attested case featuring *kayna* plus a Dutch noun.

Dutch grammatical gender has no apparent influence on MA gender assignment. Natural, biological gender, on the other hand, is a decisive factor. Besides this we may consider a number of other possible factors, in particular a) phonological shape, b) corresponding MA word, c) generalisation of either feminine or masculine as the ‘default value’. In addition, for those respondents who have received part of their education in Morocco and who also engage in MA/French CS, the gender of French cognate nouns is likely to be of influence, since French gender is often preserved in MA/French CS.

I will not attempt to investigate these possibilities in full. In view of their diverse linguistic backgrounds it is likely that the respondents have different strategies of gender assignment, but only a few respondents produced sufficient obvious examples to permit an investigation of their individual, possibly idiosyncratic system. Therefore I will confine myself to the following observations. Regarding phonological shape, MA associates word-final /a/ and, to a lesser extent, /t/ with feminine gender. In fact, most feminine nouns in MA end in /a/. This predicts that embedded Dutch words ending in /a/ are assigned feminine gender. This is indeed generally the case, as in (3), but it does not exclude nouns not ending in /a/ from assuming feminine gender as well, so that the relationship between gender and phonological shape is obscured. The grammatical influence of a corresponding MA word is always hard to establish, since we can only guess what the corresponding MA word would be. The Dutch noun *dorp* in (1), for instance, may be translated into MA as feminine *qerya* or masculine *filaž*, *dšer* or *dewwar*. The use of feminine as the default value for embedded nouns is manifest, although not without exceptions, in Samir’s speech. This tendency is less obvious amongst the other respondents. Insertion of both French and Dutch nouns is found in Zineb’s CS variety of MA. When Zineb embeds Dutch nouns with close cognates in French, the assigned gender is in accordance with that of the French word. Thus she treats Dutch *niveau* as masculine, and *administratie* and *discriminatie* as feminine (cf. French *niveau*, *administration*, *discrimination*).

5.1.2 Collective nouns

MA has a class of collective nouns which are not marked for plural, but which denote a group or class of people or objects. In Caubet’s words, “ils désignent des ensembles d’individus ou d’objets appréhendés comme un tout, comme une collection d’éléments” (1993, I: 107). A singulative form can be derived by adding the suffix *-a* in the case of non-humans, and *-i* in the case of human collectives, e.g. *nḥel*, *neḥla* “bees, a bee”; *ihud*, *ihudi* “Jews, a Jew” (cf. Durand, 1994: 102). According to the grammatical descriptions, non-human collective nouns are formally singular, although many speakers in fact oscillate between the use of singular and plural agreement patterns for non-humans.²

² This is nicely illustrated by the narratives based on Mayer’s 1969 picture book *Frog: Where are you?* which are reproduced in El Aissati (1996: 198-205). Three

There is only marginal evidence that Dutch nouns are assigned to the MA class of collective nouns. The following utterance by Abdelkrim contains a possible example of a Dutch-origin collective noun.

- (8) *ḍerḇ-eh u gaʃ knikker teyyeh-hũm u ḍerḇ-eh f straat*
 hit-3M and all marble make-fall-3PL and hit-3M in DEF-street
 “He hit him, he threw all the marbles to the ground and he hit him in the street.” (Abdelkrim)

In (8) the singular form *knikker* “marble” is remarkable. From the context we expect the plural form *knikkers* (see the next section on Number) after *gaʃ* to express “all the marbles”. The omission of a plural marker is not a recurrent phenomenon in the data corpus. *Gaʃ* plus singular noun would mean “the entire X”. The latter meaning is obviously not intended here, unless *knikker* is considered a collective noun in Abdelkrim’s system of MA noun categories. The concept of marbles, being small objects that typically occur in larger numbers, is a likely candidate to be rendered into a collective noun in MA. Finally, note that the left-dislocated NP *gaʃ knikker* is co-indexed with the plural object suffix *-hũm*. See also the discussion of *xuld ~ xulda* “guilder” below (ex. (62)-(59) in 1.4).

Regarding human collectives, we note the use of the form *surinam* to denote “Surinamese people”, instead of Dutch *Surinamer-s* “Surinamese-PL”, or Arabic *surnami-yin*. Consider the next two examples where Mustafa, a beginning learner of Dutch, and the Dutch born Nawal comment on the ethnic diversity in their classroom. (The word *ha* in (9) is a presentative particle commonly used in enumerations; for *bawtlans* in this example see the discussion in 1.10)

- (9) *ḥna-ya mxellṭ-in temmak, bawtlans mxellṭ-in: ha turk ha mḡarba*
 1PL-EMPH mixed-PL there foreigners mixed-PL PRES Turks PRES Moroccan·PL

 ha surinam ha pakistan ha kũll-ši mxellṭ-a³
 PRES Surinam(ese) PRES Pakistan(is) PRES everything mixed-F
 “We are mixed there, mixed foreigners: there are Turks, Moroccans, Surinamese, Pakistanis, there’s everything mixed.” (Mustafa)

- (10) *hadi t-turk u l-mḡarba u s-sbanyuli-yin u s-surinam*
 DEM DEF-Turk and DEF-Moroccan·PL and DEF-Spaniard-PL and DEF-Surinam(ese)

narrators (in Casablanca, Tangier, and Oujda) all produced masculine singular as well as plural verbs with *nḥel* “bees” as Subject.

³ As *kũll-ši* “everything, everyone” is normally masculine, the feminine ending in *mxellṭa* is unexpected.

u *zuid afrika* u..
 and South Africa(ns) and
 “There are Turks, Moroccans, Spaniards, Surinamese, South Africans and ..”
 (Nawal)

In the above two examples the word *surinam* is coordinated with the MA plural form *mġarba* “Moroccans” and the MA collective *turk* “Turks”, which suggests that *surinam*, too, is conceived of as a collective. On the other hand, these words are also coordinated with *pakistan* and *zuid afrika*, which are the names of these countries in Dutch. This suggests that Mustafa and Nawal simply use the name of the country to denote the people. The Republic of Surinam is known in Dutch as *Suriname*, which is pronounced with a final schwa, yet the dropping of the final schwas is a recurrent feature of Moroccans’ variety of Dutch. In any case the singulative of *surinam*, which would be *surinam-i* “(a) Surinamese” does not occur in the data (cf. *turki* “Turk”). Hayat inserts the Dutch word *Surinamer* for “a Surinamese” in MA contexts (see (119) below).

5.1.3 Number

Both MA and Dutch distinguish between singular and plural nouns. MA also has dual forms for a very restricted set of nouns, but the distinction of dual is not productive. MA plural nouns are marked either by one of several patterns of word-internal vocalic changes (ablaut or ‘broken’ plurals, e.g. *seržem* “window” pl. *sražem*), or by one of the suffixes *-at*, *-a*, and *-in*. Dutch plurals are marked by the suffixes *-en* and *-s*, the choice of which is largely phonologically determined. In addition, there are a few irregular plurals in Dutch. Embedded Dutch plurals are invariably marked by Dutch suffixes.⁴

- (11) ma ʔend-ek-š *bewijz-en*
 NEG at-2-NEG evidence-PL
 “You don’t have evidence.” (Hocine)

The agreement features in the next two examples show that Dutch plural nouns function as plurals in the MA nominal paradigm.

⁴ In other corpora plurals of the type [Dutch noun + MA *-at*] may be found, e.g. *stichting-at* “foundations” (Nortier, 1990: 180, 189). In the speech of MA/Dutch bilinguals, however, the suffix *-at* is not very productive with Dutch nouns. It should further be noted that what seem to be Dutch nouns with MA plural ending often have French cognates and exist as loanwords in Morocco, e.g. *riportaż-at* “reports”, *hurmun-at* “hormones” (Azghari, 1994: 49).

- (12) duk_i artikel-en_i, ila bġi-ti t-terżem-hūm_i, is echt moeilijk
 DEM-PL article-PL if want-2SG 2-translate-3PL is really difficult
 “Those articles, if you want to translate them, that’s really difficult.” (Hocine)
- (13) kayen serie-s_i ka-n-dir volg-en, ne-tferreż-hūm_i
 EXIST serial-PL ASP-1-do follow-INF 1-watch-3PL
 “There are TV serials that I follow, that I watch.” (Abdelkrim)

MA grammar determines whether a singular or a plural form is appropriate. Although the contexts where nouns are marked for plural largely coincide in both languages, some instances do indicate the supremacy of matrix language grammar. In (14), for instance, the use of the definite singular noun *l-buitenlander* is in accordance with MA grammar, since in generic expressions, MA may use both the definite singular and definite plural of count nouns (Caubet 1993, vol.II: 295-7). The Dutch equivalent of this utterance would require an indefinite plural. Here the absence of the EL plural marker in (14) is dictated by ML grammar.

- (14) ka-y-ṭelb-u l-buitenlander b šifa šamm-a
 ASP-3-ask-for-PL DEF-foreigner (SG) with manner general-F
 “They ask for foreigners in general.” [as opposed to foreign women only]
 (Zineb)

5.1.4 Definiteness

MA nouns are marked for definiteness by means of articles. I distinguish the definite article *l-*, and three ways of marking indefiniteness: the composite article *waħed l-*, the article *ši*, and the so-called zero article \emptyset (i.e. absence of overt marking). The use and distribution of these four ‘articles’ is rather complicated and cannot be presented here, however, we may note that *l-* has the widest distribution, followed by *waħed l-* (cf. chapter III in Caubet, 1993, vol. II). The definite article *l-* is often repeated with attributive adjectives, e.g. *l-mudun l-kbar* DEF-cities DEF-big-PL “the big cities”.

The prefix *l-*, either as the definite article or as part of the indefinite composite article *waħed l-* assimilates to the initial consonant of the noun or adjective if this is a coronal other than [y]. Assimilation does not always apply to the voiced alveo-palatal fricative /ʒ/ [3]. Assimilation results in a so-called geminate consonant, e.g. *r-riħ* “the wind”. Geminate consonants constitute a serious and often underestimated complication for the analysis of the data, because a consistent and reliable distinction of geminates and non-geminates is not so evident. In the context of MA/Dutch CS the problem of geminate consonants is particularly relevant for the MA definite prefix, which is the only MA affix that is productive with embedded Dutch stems. Therefore the problem needs to be clarified here.

Geminate consonants

The concept of geminate consonants is indispensable for an understanding and analysis of MA morphology. Geminata sometimes leads to resyllabification, as in the verbal form *xerrež* “he took out; made go out” (2 syllables) which is the causative of *xrež* “he went out” (1 syllable). Some morphological processes also separate geminate consonants by inserting a vowel. So for example the plural noun *dfuf* “(kind of) tambourines” shows that the singular *deff* ends in a geminate /f/, at least underlyingly.

In perception, MA geminate consonants are sometimes difficult to distinguish from non-geminates, at least for non-native speakers. Doubts arise particularly concerning geminates in the onset or coda of a syllable. Keegan (1984: 49; 1986: 29) maintains that MA geminates are reduced in all syllable final positions, which he illustrates by such alternations as *sem* “paternal uncle” versus *sem-i* “my uncle” and *xed* “cheek” versus *xdud* “cheeks”. On the other hand, Heath makes no mention of such degemination in his 1987 book on MA phonology.

With respect to the syllable onset, we note the differing representations of the medio-passive and reciprocal prefix. According to Heath (1987: 77, 280-4), this prefix has the simple and geminate allomorphs /t-/ and /tt-/: /t-/ ‘typically’ occurs before stems beginning in /CV../ as in *t-baš* “it was sold”; /tt-/ is ‘typical’ before /CCV../, as in *tt-qtel* “he was killed” (likewise Caubet, 1993, I: 33).⁵ Youssi (1992: 80) and El Aissati (1996: 111), on the other hand, claim that the prefix has the geminate form /tt-/ in all contexts regardless of syllable structure, e.g. *tt-kal* “it was eaten”, *tt-katb-u* “they wrote each other”. In Durand (1994: 68, 79) the prefix is /tt-/ for underived (so-called form I) verbs, regardless of syllable structure, and /t-/ for derived verbs, thus *tt-kal* “it was eaten”, but *t-katb-u* “they wrote each other”. The confusion that reigns over the representation of this prefix suggests that its perception as a simple or geminate consonant is influenced by the perceiver’s conception of MA syllable structure in general. Unless, of course, the divergent analyses can all be ascribed to idiolectal or dialectal variation (Harrell, 1962: 33).

A perception experiment carried out by Obrecht (1965) involving Lebanese Arabic stimuli further suggests that geminates in onset position are not as easily distinguishable as those in intervocalic position.⁶ The transcription of MA by native

⁵ According to Heath, the allomorph /tt-/ also occurs before /CCCV../ verbs. He applies this latter pattern to forms such as *xlleš* “he payed” and *xbba* “he hid”, which other authors prefer to represent with an epenthetic schwa (*xelles*, *xebba*) thus ranging them under the /CV../ type of verbs.

⁶ For Lebanese Arabic, some rather small-scale experiments were carried out by Obrecht (1965) and Miller (1987). Miller’s production experiment with one respondent shows that intervocalic geminates are longer (approximately 2.5 times the length of the single counterpart) than geminates in onset or coda position (approximately 2.0 times the length of the single counterpart) (1987: 136). In Obrecht’s perception experiment involving synthesized stimuli even a 2.5 times longer initial

speaker linguists casts some further doubts on the realisation and/or perception of geminate consonants. Non-native speakers tend to rely entirely on morphological considerations in their transcription of geminates. But in the transcripts of native speakers, who can be expected to rely more on their auditive impressions, the distribution of geminates is seldom consistent with the distribution based on morphological analysis.⁷ It may be that geminate consonants in some contexts are a psychological rather than an acoustic reality, and that they are reduced in the pronunciation. Therefore I believe that the discrimination of geminates as opposed to single consonants by native speakers of MA needs to be investigated by means of production and perception experiments. In the absence of definite information regarding the perception of the different MA geminates in various phonological contexts, a reliable discrimination between the single and geminate consonants was not feasible within the framework of this study.

Because my perception of geminate consonants is insufficiently reliable, I will only consider the presence and absence of the definite prefix in contexts where it is not part of a geminate cluster. This largely excludes a discussion of the prefixation to Dutch nouns (and adjectives) that begin with a coronal consonant (other than [y]), unless the prefix appears in its unassimilated form (see ex. (21) and (22)). Thus no attempt will be made to decide whether or not in (15), for example, the Dutch word *scheikunde* contains the assimilated MA prefix *l-* (*s-scheikunde*).

consonant in the minimal pair *šábiy* “boy” and *š-šábiy* “the boy” was not unequivocally perceived as a geminate: “while an 80 msec. noise duration was identified by these three subjects as 100% /š/, a value as high as 200 msec. was judged to represent /šš/ only 70.37% of the time; this despite the fact that spectrographic measurements had indicated that the range from 80 msec. to 200 msec. noise duration should have been adequate to cue this distinction with more-or-less 100% accuracy, if duration were the only cue involved in the discrimination” (Obrecht, 1965: 38-9). Obrecht’s other experiments, with the minimal pairs *xábar* “news” and *xábbar* “he informed”, and *bána* “he built” and *bánna* “mason”, showed far less confusion over single and geminate consonants. The experimental data suggest that geminates are better perceived in intervocalic position, although the author does not consider this explanation.

⁷ So Moroccan linguists tend to transcribe for instance *weždat* “she prepared” (phonetically correct?) instead of *wežždat* (morphologically correct) (El Idrissi, 1990: 40), or to alternate between both spellings. Cf. also Bentahila & Davies (1991) *l-muhima* (p. 388), *ka-tfeḍli*, *l-ḥuriya* (p. 389) instead of *l-muhimma*, *ka-tfeḍḍli*, *l-ḥurriya*; Nait M’barek & Sankoff 1983 *mxalšin* (p. 145) *ta-ybadlu* (p. 148) instead of *mxellšin*, *ta-ybeddlu*. Note that these are examples of a phenomenon that is pervasive in nearly all native speakers’ transcripts.

- (15) ma ʕli-h-š waš ne-qra *scheikunde* wella ne-qra fraṇṣawiya wella
 NEG on-3M-NEG Q 1-study chemistry or 1-study French or
 l-inglesiya
 DEF-English
 “It doesn’t matter whether I study chemistry, or French, or English.”
 (Samir)

Moreover, we must exclude from the analysis those cases where the definite prefix, if present, must be assumed to form a geminate cluster with the final [l] of the preceding word. This particularly concerns the MA prepositions *dya* “of”, *bħal* “like”, and the clitic prepositions *l* “to, for” and *fel* “on, over” (allomorph of *lla*). So in (16) the preposition *bħal* makes the perception of the definite prefix unreliable. In (17) it is uncertain whether it is the preposition *l* that is realised, or the definite prefix *l-*, or both (both the preposition and the article tend to be omitted, as we will see below).

- (16) ġadi ye-wqeʃ (..) bħal [l-]getto lli kan ʕend-hũm l-ihud hna-ya
 FUT 3-happen like [DEF-]ghetto REL be at-3PL DEF-Jews here-EMPH
 “It will happen (..) like the ghetto the Jews used to have here.” (Maryam)
- (17) walakin te-mši matalen l *godsdiens*t wella l ši ħaža
 but 2-go for-example to/DEF-religion or to INDEF thing
 “But if you look at the religion or something, (..)” (Hocine)

If we disregard the contexts where gemination obscures the perception of the prefix *l-*, it becomes apparent that embedded Dutch nouns occur with all four ‘articles’. Examples of each article are reproduced below.

i. *The definite article l-*

Although this article tends to be omitted, as I will discuss presently, there are also many instances where it does surface before Dutch nouns.

- (18) wah, la, maši dar l-goal, b š-šehħ kan y-dir l-kura u beʕda (..)
 yes no NEG do DEF-goal with DEF-truth be 3-do DEF-ball and then
 “Yeah. No, he didn’t make the goal, but he manipulated the ball and then ..”
 (Nawal)
- (19) ta l-waarheid lqi-na-h f ž-ždida f l-lycée *technique*
 even DEF-truth find-1PL-3M in El Jadida in DEF-school technical (French)
 “Even the truth, we found it in El Jadida, in technical school.” (Mimoun)
- (20) ana, š-šaraħa, ik heb geen behoefte aan hē, ġir l-weekend, alleen
 1SG DEF-frankness I have no need for QT only DEF-weekend only

maar in het weekend

just in the weekend

“Frankly, I’ve no need for it [i.e. to go out], right? Just the weekend, only in the weekend.” (Jamal)

Occasionally, the unassimilated form of the prefix is found before a Dutch noun beginning with a coronal consonant (other than [l]), as in the following example (see also (22)).

- (21) had n-nas, l-regering t-redd-hūm l blad-hūm, te-ŷti-hūm
DEM DEF-people DEF-government 3F-send-back-3PL to country-3PL 3F-give-3PL

l-flus lli ye-tqedd-u

DEF-money-PL REL 3-suffice-PL

[If the Dutch decide to deport foreign workers,] “These people, the government should send them back to their country and give them sufficient money.” (Hayat)

ii. *The indefinite composite article waḥed l-*

Preceding a Dutch noun, the composite article *waḥed l-* usually takes the form *waḥed*, omitting the prefix *l-*, as I will discuss below. The only attested examples *waḥed l-* plus Dutch noun occur in the speech of Samir.⁸

- (22) ka-t-gul waḥed l-żeww, waḥed l-sfeer lli ka-ye-xleq
ASP-2-say INDEF DEF-atmosphere INDEF DEF-atmosphere REL ASP-3-create

kūll-ši?

everything

“You say [it’s] a certain atmosphere that creates everything?” (Samir)

- (23) plato ta huwa ŷend-u waḥed theorie, maši theorie, waḥed l-.. verhaal
Plato also 3M at-3M INDEF theory NEG theory INDEF DEF- story

ŷla waḥed ř-razel huwa mkettef f waḥed l-.. l- eh grot

about INDEF DEF-man 3M chained in INDEF DEF-.. DEF- er cave

Plato also has a theory, not a theory, a .. story, about a man who is chained in a er .. cave. (Samir)

⁸ Other examples probably concern French cognate words rather than Dutch nouns. Mimoun for instance provides *waḥed l-interview* “an interview”.

Note that in (22) the prefix *l-* is not assimilated to the initial coronal consonant of Dutch *sfeer*, “atmosphere”. Lack of assimilation is not a recurrent feature of Samir’s speech. It is fairly common for the definite prefix not to assimilate to the voiced alveo-palatal fricative, as in *l-žeww* in the same example. In (23) we see that the occurrence of *waḥed l-* preceding a Dutch noun is related to a hesitation phenomenon in which the string *waḥed l-* is produced as a phonological word, followed by a pause. I will return to this later on.

iii. *The indefinite article ši*

The indefinite *ši* is freely used with embedded Dutch nouns.

- (24) *ila kūn-t ġadi ne-šri ši lijst, ġadi hadik hiyya l-lewwl-a lli ġadi*
 if be-1SG FUT 1-buy INDEF frame FUT DEM 3F DEF-first-F REL FUT

n-dir-ha l-cadre

1-do-3F DEF-frame (French)

“If I’m ever going to buy a frame, that one will be the first that I frame.”

(Mimoun)

- (25) *ne-mši-w l ši friettent wella la?*
 1-go-PL to INDEF snack-bar or NEG
 “Shall we go to a snack bar, or not?” (Jamal)

- (26) *ʔend-ek ši hobby-s*
 at-2 INDEF hobby-PL
 “Do you have any hobbies?” (Samir)

iv. *The indefinite zero article*

The zero article must not be confused with those cases where the definite prefix fails to surface. One of the contexts where the zero article is to be expected is the predicate in copula constructions (Caubet, 1993, II: 260).

- (27) *f had l-marḥala ka-y-welli communist*
 in DEM DEF-stage ASP-3-become communist
 “At this stage one becomes a communist.” (Najib)
- (28) *ʔla l-ʔamal ʔanna-ni n-welli lerares f l-musteqbel*
 on DEF-hope CONJ-1SG 1-become woman-teacher in DEF-future
 “In the hope that I will become a teacher in the future.” (Zineb)

5.1.5 Omission of the definite prefix

The omission of the MA definite article before foreign stems has been noted by other scholars with respect to loanwords from Berber and Hispanic Romance (Colin, 1945: 232; Harrell, 1962: 190). In MA/Dutch CS, the attachment of the prefix *l-* to embedded Dutch nouns is variable, yet all respondents tend to omit it in contexts where it is obligatory in MA. Note that the definite prefix *l-* is the only MA affix that is at all productive with Dutch (content) morphemes.

The omission concerns both the definite article *l-* and *l-* as part of the indefinite composite article *waḥed l-* or the demonstrative forms *had l-* (proximate) and {*dak*, *dik*, *duk*} *l-* (remote M, F, PL). Note that in the indefinite and demonstrative contexts the prefix *l-* is always required according to MA grammar. This also means that the prefix is redundant, since there is no difference in meaning between *waḥed* and *waḥed l-*, or *had* and *had l-*. I will examine the omission phenomenon in these contexts separately. In the examples cited the underlined space indicates the missing *l-* prefix.

i. Omission of *l-* in the indefinite composite article *waḥed l-*

After *waḥed* the omission of *l-* is almost consistent, see the discussion of ex. (22) above. Two examples are reproduced here; see also (2) above and (40) and (70) below for further examples.

- (29) u ʃad te-bqa waḥed -*probleem* dyał ‘je moet zoveel diploma-s
and still 3F-remain INDEF problem of you must so-many diploma-PL

hebben om dat te doen

have for this to do

“And there’ll always remain a problem of ‘you need so many diplomas to do this’.” (Fatima)

- (30) ka-n-dewwez waḥed -*cursus*
ASP-1-follow INDEF course
“I follow a course.” (Mustafa)

ii. Omission of *l-* in the demonstrative forms

After the use of one of the MA demonstrative determiners, the definite prefix is consistently omitted by most of the respondents. This is exemplified in (31), and also in (12) above.

- (31) ka-n-tebbeʃ dak ehm -*opleiding* dyał *leraar-opleiding*
ASP-1-follow DEM er training of teacher-training
“I follow that training, the teacher training.” (Zineb)

Only Hayat and Samir show variation in this respect: the definite prefix surfaces in examples (32) and (34), whereas it is lacking in (33) and (35).

- (32) *ža-w huma b dik l-idee men temma l hna*
 come-PL 3PL with DEM DEF-idea from there to here
 “They brought this idea from there to here.” (Hayat)

- (33) *šnu had l-xūḍṛa, ki smiyt had _-groente*
 what DEM DEF-vegetable how name DEM vegetable
 “What’s this vegetable, what’s the name of this vegetable?” (Hayat)

- (34) *u hnaya had l-arbeider-s dyal l-huḷandī-yin huma ta huma ſend-hūm*
 and here DEM DEF-worker-PL of DEF-Dutch-PL 3PL also 3PL at-3PL

l-ʔafkar ḍeyyq-in, ḍeyyq-a

DEF-idea·PL narrow-PL narrow-F

“And here these Dutch workers they also have narrow ideas.” (Samir)

- (35) *l-walid-in dyal-ek ma ka-y-fehm-u-š eh had eh _-wereld-je*
 DEF-parent-PL of-2SG NEG ASP-3-understand-PL-NEG er DEM e r
 world-DIM

(waar-in je leeft)

where-in you live

“Your parents don’t understand this small world in which you live.” (Samir)

iii. Omission of the definite prefix in other contexts

In various other contexts the missing definite prefix can be clearly distinguished from the ‘zero article’: in (36) below the definite article would be required because the noun *cultuur* is determined by the analytic possessive *dyal-ek* “your”. Then in (37) and (38) we would expect the definite prefix to precede the EL words *gevoel* and *boeddisme* because these are non-count nouns (Caubet, 1993, II: 274)⁹; moreover, *gevoel* in (37) is specified by the complement clause. The definite prefix can also be considered lacking in the above examples (3), (6), (7) and (8), for instance.

⁹ Hocine pronounces the EL noun as [budism], dropping the final schwa of the Dutch target word. The dropping of the word final schwa is a common feature of Moroccan learners’ varieties of Dutch. On the other hand, the word can be understood as the French cognate *bouddisme* [budism]. In this case, however, the omission of the definite article diverges from what is known from earlier studies on MA/French CS, where embedded French masculine nouns receive either the MA definite prefix or the French definite article *le*.

- (36) ta-te-qra-y _-cultuur dya-ek u d-din dya-ek
 ASP-2-learn-F culture of-2SG and DEF-religion of-2SG
 “You learn about your culture and your religion.” (Hayat)
- (37) ma ʕend-hūm-š _-gevoel baš y-ʕerf-u l-ḥlawiya dya l-franšawiya
 NEG at-3PL-NEG feel COMP 3-know-PL DEF-sweetness of DEF-French
 “They don’t have the feel to know the charm of the French language.”
 (Najib)
- (38) huma žab-u-hūm men _-boeddhisme
 3PL get-PL-3PL from Buddhism
 “They got them from Buddhism.” (Hocine)

In obligatory contexts other than after *waḥed* or a demonstrative, the data corpus shows much variation with regard to the application of the MA definite prefix to embedded Dutch nouns, although omission occurs in the speech of all respondents. The morpho-phonological context is a factor that influences the omission phenomenon; some tentative observations concerning this appear below.

Realisation of the definite prefix in certain morpho-phonological contexts

With respect to ex. (23) above, I pointed out that the definite prefix can be uttered independently from the head noun when the speaker has trouble finding the appropriate word. This is particularly so if the prefix follows the indefinite *waḥed*, the clitic preposition *f*(locative), or the demonstrative *had*. In the case of hesitations these strings form the phonological words [wəḥdəl], [fəl] and [hədəl] (in which the epenthetic central vowel can be very short). In such hesitation contexts juxtaposition of the prefix *l-* and a Dutch noun does occur, but a pause intervenes between the two, so that *l-* is not actually realised as a prefix. The hesitation factor is clearly demonstrated in the next example where Samir’s favourite filler word *ngulu* “let’s say” comes between the composite indefinite marker and the embedded Dutch noun.

- (39) ana ʕend-i waḥed l- eh n-gul-u gewoonte baš n-..
 1SG at-1SG INDEF DEF er 1-say-PL practice COMP 1-..
 “I have a, let’s say, practice to ..” (Samir)

Examples of *waḥed l-* [wəḥdəl] plus a Dutch noun as a hesitation phenomenon were only found in Samir’s speech.¹⁰ Apart from this, note that it is not uncommon in

¹⁰ This is probably explained by the fact that his contributions to the conversations make up a relatively large part of the recorded material, while, at the same time, hesitation is an apparent characteristic of his speech in MA or MA/Dutch CS.

(monolingual) MA for *waḥed l-* to be uttered separately from the noun. The following example from Mimoun illustrates this.

- (40) ma kayen ši ʔinsan, ʔinsan, ʔinsan lli ma ʃend-u-š waḥed,
 NEG EXIST NEG human human human·being REL NEG at-3M-NEG INDEF

waḥed l- eh, waḥed l-merḥala, maši merḥala wa-ʔinnama waḥed
 INDEF DEF er INDEF DEF-stage NEG stage and-but INDEF

l-leḥḍa, waḥed eh *-toestand* lli ka-y-welli fi-ha ḥayawan
 DEF-moment INDEF er situation REL ASP-3-become in-3F beast

“There’s no human, human, human being who doesn’t have an er .. a stage, not a stage but rather a moment, a situation in which he becomes a beast.”
 (Mimoun)

Also apparent in Samir’s speech is the tendency to realise the definite prefix after the clitic preposition *f* “in”, as in (41)-(44). Counter-examples also occur, however, cf. (45). Example (41) actually shows the insertion of an adjective-noun collocation; this type of insertion will be addressed later on.

- (41) wašta ka-t-dir f l- eh f l-*vrij-e* *tijd*
 what ASP-2-do in DEF- er in DEF-free-AGR time
 “What do you do in your spare time?” (Samir)

- (42) ah, maši bḥal l-mḡarba lli ka-ne-ʃref-hūm ana f l- .. eh
 oh NEG like DEF-Moroccan·PL REL ASP-1-know-3PL 1SG in DEF- er

omgeving dya*-i*
 environment of-1SG

“Oh it’s not at all like the Moroccans I know in my er environment.”
 (Samir)

- (43) kan y-dir-hūm hnaya, zeʃma f l- eh *schoot* dya*-u*
 be 3-put-3PL here EPIST in DEF- er lap of-3M
 “He used to put them here, let’s say, on his lap.” (Samir)

- (44) t-geʃd-i f l-*bank*, ya-k?
 2-sit-F in DEF-couch QT-2
 “You sit on the couch, don’t you?” (Samir)

- (45) waš t-kun-i f *kan..* f *kantine* wella ..
 Q 2-be-F in can.. in canteen or
 “Will you be in the canteen or ..?” (Samir)

Due to the speaker's hesitation, the syllable *fl-* [fəɫ] “in the” is uttered as a separate word in (41)-(43) (but not in (44)). This also explains the non-assimilation of *l-* to the initial [s] of *schoot* [sXot] in (43). The same phenomenon obtains when the definite prefix merges with the preposition *men* “form, out of”, so that *men l-* “out of the” is realised as [məɫ], which is a common assimilation found in MA.

- (46) ta nta-ya *links*, siwa nta-ya ši xetɾ-at ka-t-t-ğeyyer men
 too 2M-EMPH left except 2M-EMPH INDEF time-PL ASP-2-MP-change from

 m l- eh m l- *kleur*
 from DEF er from DEF colour
 “You are leftist too, but you just sometimes change colour [i.e. political persuasion].” (Samir)

When the definite prefix is part of the demonstrative construction *had l-*, we observe a similar phenomenon in Samir's speech. Likewise, the realisation of *l-* in *be ſd l-h.. handelingen* “some actions” in (47) can be related to the fact that the quantifier *be ſd* is always followed by the definite prefix, with the two morphemes together constituting a phonological word [bəſdəɫ].

The main facts concerning the productivity of the MA definite prefix with Dutch nouns can be summarised as follows. After the indefiniteness marker *wahed* the prefix is (almost) consistently absent; after one of the attributive demonstratives the prefix is absent in the large majority of the cases, however the speech of two respondents shows variation in this respect. In the remaining obligatory contexts, where *l-* is the only determiner, its presence is highly variable, omission being frequent in the speech of all respondents.

5.1.6 Quantification

MA expressions for quantification can be classified into a number of syntactic structure types. One type involves quantification words that are formally equivalent to attributive adjectives, in that they follow the quantified noun, and agree in gender and number, e.g. *n-nas kaml-in* [DEF-people entire-PL] “all the people”. The other major type is formally similar to what I will call genitive constructions, in which the quantifier precedes the quantified item. There are two types of genitive in MA, one ‘synthetic’ and the other ‘analytic’. In the synthetic construction the quantifier immediately precedes the quantified NP; in the analytic genitive one of the particles *dyal*, *d-* or *nta ſ* links the quantifier to the quantified NP. I will comment further on genitive constructions in section 1.5 on possession.

There are several examples of MA numerals or other quantifiers with Dutch nouns. Firstly, concerning other types of quantification constructions than numerals, viz. expressions for “some”, “every” and “all”, (47)-(48) show examples of the synthetic genitive (cf. also *ga ſ knikker* in (8)).

- (47) ʃla xaṭeɣ ʃend-ek ʃi beʃḍ l-h.. *handeling-en* (*die je doet* (..))
 because at-2SG INDEF part DEF-h.. action-PL that you do
 “Because you have some .. some actions that you do (..).” (Samir)
- (48) xesʃ-ek f kull eh *wijk* wella kull derb y-kun-u sakn-in ʒuʒ, tlata,
 must-2SG in every er district or every alley 3-be-PL live-PL two three
 baraka!
 that’s all
 “In every district or every alley there should be two or three [immigrants], no more.” (Hayat)
- (49) der-t had l-ʃam, had eh *eerste semester*, gaʃ haduk __-*vakk-en*
 do-1SG DEM DEF-year DEM er first semester all DEM subject-PL
 “I finished this year, this er first semester, all those subjects.” (Samir)

While the synthetic genitive construction is restricted to a small number of fixed expressions, the analytic genitive is found in all kinds of less idiomatic types of quantification. The latter construction is used, for instance, with measures of capacity for mass nouns like “coffee” or with percentages:

- (50) n-dir *lekker-e* eh dak eh kas kbir-a, dak l-kas kbir-a dyal *koffie* u
 I-make nice-AGR er DEM er glass big-F DEM DEF-glass big-F of coffee and
 ne-qra l-*krant*, (...)
 I-read DEF-newspaper
 “I make a nice er such er big mug, such a big mug of coffee and I read the newspaper ..” (Jamal)
- (51) xemsin f l-miya wella ʔaktaɣ dyal eh *analfabet-en*
 fifty in DEF-hundred or more of er illiterate-PL
 “Fifty percent or more of illiterates” (Najib)

Numerals

Numerals tend to occur in the same language as the counted noun (cf. 4.3 below). However, some examples of MA numbers with Dutch nouns appear in the data, mostly in the contributions of Samir and his younger siblings, Nawal and Abdelkrim. All attested examples are reproduced here.

Counting is rather complicated in MA. Disregarding a small class of nouns with dual forms, the system is as follows. Where the number “one” is emphasized, the word *wahed*, feminine *wahda* “one” is placed after the noun like an attributive adjective, thus contrasting with the invariable and prenominal indefinite article.

Moreover, while the indefinite article *waḥed* requires the definite prefix *l-*, the numeral precludes it. Hayat provides a ‘CS example’ of the numeral *waḥed*:

- (52) *la, huma ka-y-šift-u miyat turki u meğrabi f wijk waḥd-a!*
 no 3PL ASP-3-send-PL hundred Turk and Moroccan in district one-F
 “No, they send a hundred Turks and Moroccans to one district!” (Hayat)

In connection with the use of the number “two”, two constructions are commonly encountered. One resembles the synthetic genitive: the number *žuž* is juxtaposed to the PLURAL noun, without an intervening definite prefix [*žuž* N·PL]. The other construction is similar to the analytic genitive, with an intervening preposition *dyal* or *d-*, and the definite prefix, [*žuž dyal l-N·PL*]. This latter construction is illustrated in the following example from Najib, where the Dutch noun is in apposition with its MA translation equivalent.

- (53) *kber-ti bin žuž dyal eh smiyt-u, t-tafaq-at eh cultur-en*
 grow-2SG between two of er name-3M DEF-culture-PL er culture-PL
 “You grew up between two er what’s it, cultures.” (Najib)

For the numbers 3 to 10, MA uses the analytic construction with the prefixed PLURAL noun: [{3, 4, .. 10} *dyal l-N·PL*]. Then, from 11 to 19, some dialects have the same analytic construction, but most often the synthetic genitive construction is used with the prefix-less SINGULAR noun, i.e., [{11, .. 19} N·SG]. The latter construction for the numbers 11 to 19 is probably part of the dialect of the Hamadi siblings. Finally, all speakers use the latter construction with numbers from 20 onward: [{20, 21, ..} N·SG], cf. *miyat turki u meğrabi* “hundred Turks and Moroccans” in (52) above.

This complex MA counting system is somewhat confused in the Hamadi siblings’ variety of MA, at least when EL Dutch nouns are counted, as the following examples show. First consider the number “one” in the following passage, where Abdelkrim stresses that he had only one unsatisfactory mark at school. The emphasised number *waḥed* precedes the noun instead of following it like an adjective (compare (52)).

- (54) *žeb-t i waḥed onvoldoende, (u men beṣd hadikl-xetṛa ma*
 take-1SG only one unsatisfactory·mark and from after DEM DEF-time NEG

žeb-t-š gaṣ onvoldoende; žeb-t i zess-en of zeven-s of
 take-1SG at·all unsatisfactory·mark take-1SG only six-PL or seven-PL or

acht-en-s)

eight-PL-PL

“I got just one unsatisfactory mark, and after that time I didn’t get any unsatisfactory marks at all; I got only sixes, or sevens, or eights.” (Abdelkrim)

In Samir's utterances, the number *žuž* "two" occurs several times with a Dutch noun, as illustrated by the next two examples.

- (55) u [ž-]žuž *journalist-en* lli ka-neʃref-hūm, ta huma mša-w f *ħakk*
and [DEF?]-two journalist-PL REL ASP-1-know-3PL also 3PL go-3PL in dem

š-šher

DEF-month

"And the two journalists I know, they were also gone that month." (Samir)

- (56) kūn-t ne-ħdi waħed d-ḍar dyal žuž eh žuž dyal *journalist-en*
be-1SG I-guard INDEF DEF-house of two er two of journalist-PL
"I was looking after the house of two er two journalists." (Samir)

The above examples show that Samir alternates between the synthetic and the analytic genitive construction for the number "two". This parallels the variation in Samir's monolingual MA stretches, e.g. *žuž žwame* "two mosques", *žuž de-l-masa ʃil* "two matters". As earlier mentioned, both these constructions occur in MA, except that in (56) the definite prefix may be missing.¹¹ Samir's oscillation between both constructions may result from his contacts with speakers of other MA dialects in the Netherlands.

For the remaining numbers, we have just one example from Nawal, and another from Abdelkrim:

- (57) ʃla l-ʔaqell ana ne-t-berrek ši xemsa.. ʃešra *minuut*
on DEF-least 1sg I-MP-make·sit(?) INDEF five ten minute
"I lie on for at least some five, ten minutes." (Abdelkrim)

- (58) ne-tsenna ši eh *kwartier*, xemstaš eh *minuut* .. *minut-en*
I-wait INDEF er quarter·hour fifteen er minute minute-PL
"I wait about a quarter of an hour, fifteen minutes." (Nawal)

In (57) we encounter the indefinite singular noun *minuut* in place of the expected preposition *d* ~ *dyal* plus the definite plural noun, after the number "ten". The indefinite singular is correct after the number "fifteen" in (58), but here we notice Nawal's indecisiveness in selecting the correct form.

¹¹ This relates to the problematic perception of geminate consonants. In this particular example, the definite prefix could be assimilated to the initial [ʒ] of *journalisten* [ʒurnalɪstə], hence *ž-journalisten*. However, the definite prefix sometimes fails to assimilate to this consonant, particularly in the case of loanwords. In case of non-assimilation we get *dyal l-journalisten*, which is also difficult to perceive because the prefix *l-* will merge with the final [l] of *dyal*.

Guilders

Finally, we note some occurrences of the Dutch currency unit, *guilder* [XYldə(n)] “guilder”. This word is most often used in its phonologically assimilated form *xuld* or *xulda* [Xuld(a)] in MA contexts, even by respondents of the second generation who insert other Dutch words without phonological assimilation.¹² The counting system as explained above is not consistently applied to the currency unit. Perhaps the following examples are best explained if both variant forms *xulda* and *xuld* are considered mass or collective nouns. Even then, the analytic construction used by Hocine seems anomalous (cf. “hundred” in (52) above); the expression used by Zineb in (62), *mitin f l-xuld* should perhaps be understood as “two hundred in terms of guilders”.

- (59) *ila der-ti mša-na mša had l-beht ne-ſti-l-ek l-axūr eh ja, jouw eh*
 if do-2SG with-1PL with DEM DEF-research l-give-to-2SG DEF-other er yes your er

vergoeding, rebſin xulda

reward forty guilder

“If you participate with us in this research, I’ll give you, what’s it, er your er reward, forty guilders.”¹³ (Samir)

- (60) *t-xelleſ-l-hūm xemsa d l-xuld l s-saša*
 MP-pay-to-3PL five of DEF-guilder for DEF-hour
 “They were paid five guilders an hour.” (Warda)

- (61) *ka-ye-ſti-w-l-ek miya d l-xulda, d l-xulda, t-ſiſ wella la t-ſiſ*
 ASP-3-give-to-2SG hundred of DEF-guilder of DEF-guilder 2-live or NEG 2-live
 “They give you one hundred guilders, whether you live [off this] or not.”
 (Hocine)

- (62) *ka-y-axūd rebſataſel ʔalef ryal f š-šher, hiya mit-in f*
 ASP-3-take fourteen thousand ryal in DEF-month 3F hundred-DUAL in

l-xuld èh?

DEF-guilder QT

¹² In this context note that Hayat uses the word *tawzent* ~ *tawez* “thousand” when referring to sums of money. Heath (1989: 315) signals *tawzen*, for which he assumes an English or German origin, as an uncommon slang expression for “money” in Fez. The Dutch cognate is *duizend* [dœyʒənt].

¹³ Samir’s MA expression for “to participate with so. in sth.” is a clear calque of the Dutch expression *meedoen met*. In monolingual MA, the expected phrasing would be something like *ila t-šarek-ti mša-na f had l-beht* [if MP-participate-2SG with-1PL in DEM DEF-research].

“He earns fourteen thousand riyal a month, that’s two hundred guilders, right?” (Zineb)

5.1.7 Possessive marking

Possession is a kind of noun modification in the sense that the possessor determines and identifies the possessed noun; indeed, factual possession is implied in only a small percentage of all occurrences of possessive constructions. Embedded nouns can be marked for pronominal or lexical possessors in either MA or Dutch.

Dutch possessives

Dutch nouns with Dutch pronominal possessives of the type *zijn idee* “his idea” are considered as (subtypes of) embedded NP constituents. They form the topic of section 4.2 below. Embedded Dutch nouns modified by an analytic possessive construction in Dutch, like *vrijheid van individu* in (63), are not considered to be EL constituents. This is because this string does not constitute a possible nominal constituent in Dutch since the first noun always requires a determiner (in this particular case the absence of a determiner before *individu* is also ungrammatical). Instead, in this example, *vrijheid* is an EL noun and *van individu* an EL prepositional constituent.

- (63) qbila hḍer-t-l-ek ʕla *individueel vrijheid, vrijheid van individu*,
 previously talk-1SG-to-2SG about individual freedom freedom of individual

yeʕni l-ḥurriya l-ferdiya
 that-is DEF-freedom DEF-individual

“Before I talked to you about individual freedom, the freedom of the individual, that is [in Arabic] individual freedom.” (Hocine)

Since this example of a Dutch noun modified by a Dutch analytic possessive construction is unique in the data, the rest of this section will concern MA possessives.

MA possessives

As noted earlier, there are two types of possessive marking in MA, one synthetic and the other analytic. The synthetic construction involves either the attachment of a pronominal suffix that indicates the possessor, e.g. *blad-hūm* [country-3PL] “their country” in (21), or the juxtaposition of the possessed noun to the possessor NP, as in *smiyt had groente* [name DEM vegetable] “the name of this vegetable” in (33). The synthetic construction in MA is restricted to a small set of nouns that can occur in the possessed (modified) position. This set varies from speaker to speaker (cf. Boumans, 1994), but is confined to a relatively small number of nouns in a few semantic domains, often characterized as ‘inalienable’, in addition to fixed

expressions. As a result, it comes as no surprise that no embedded Dutch nouns were found to be modified by the synthetic construction.

The analytic possessive makes use of one of the particles *dyal*, *d* or *nta* *ʃ* to link the possessed object to the possessor. Only *dyal* and *nta* *ʃ*, which are dialectal variants of each other, are used with pronominal suffixes. These particles may be considered as prepositions, although in some dialects *dyal* or *nta* *ʃ* have allomorphs that agree with the possessed noun in gender and number, thus differing from other ('real') MA prepositions. Unlike the synthetic genitive construction, this analytic construction is entirely productive in MA, without any restrictions on the possessed noun or the possessor.

Embedded Dutch nouns are also freely combined with the MA analytic possessive construction. This is amply illustrated in the above examples (7), (29), (70), (34), (31), (36), (42) and (43). Note that in this construction, both the possessed (modified) element and the possessor can be in Dutch and the latter can even be a clause, as in (29).

5.1.8 Adjunct PPs

Like MA nouns, embedded Dutch nouns can be modified by various prepositional adjuncts (for complement PPs, see section 1.9 hereafter). The following examples show MA PPs: a locative in (64) and (65) and what we may call an instrumental adjunct in (66). In the latter case, the PP *b l- ʃerbiya* "in Arabic" can be interpreted as an adjunct to either the preceding noun or the verb phrase as a whole.

- (64) u ka-n-.. ne-qra l-ʔaxbar ʃal eh zeʃma ki semmi-w-ha,
and ASP-1-.. 1-read DEF-news·PL about er EPIST how call- PL-3F

ʃla had *toestand-en* f š-šerq l-ʔawsaʔ u ʃafi
about DEM situation-PL in DEF-East DEF-Middle and enough
"And I read the news about er what's it, about the situation in the Middle East, that's all." (Samir)

- (65) ʃta-w-ni *vrijstelling* f *vrij-e* *ruimt*[e]
give-PL-1SG exemption in free-AGR space
"They gave me an exemption in the 'free space' [un-programmed part of university curriculum]." (Zineb)

- (66) ǧadi n-dir-u hakda u n-dir-u *passage* b l-ʃerbiya
FUT 1-do-PL like·this and 1-do-PL passage with DEF-Arabic
"We'll do [it] like this, and we'll do a passage in Arabic." (Hocine)

Only one example was found of an embedded Dutch noun modified by a Dutch adjunct PP:

- (67) u hadak š-ši y-kun *nadeel* voor de universiteit?
 and DEM DEF-INDEF 3-be disadvantage for the university
 “And this will be a disadvantage for the university?” (Hocine)

5.1.9 Noun complements

Both in MA and in Dutch, nouns may subcategorize for a complement introduced by a preposition, or a complement clause. Both types of complementation are found with embedded Dutch nouns. Typically, the noun plus complement can be paraphrased by a conceptually and/or etymologically related verb that takes a clausal, prepositional, or Direct Object complement.

The complement can be in either Dutch or MA. In the former case, the Dutch complement is considered an embedded constituent in itself (clause or PP); the noun plus complement is not considered a Dutch constituent, unless the string has the distribution of an NP constituent in Dutch. In connection with this it must be remembered that embedded nouns without a determiner, although viewed as possible constituent types in Dutch within certain contexts, do not constitute sufficient evidence for constituent insertion and will be counted as embedded content words.

MA prepositional complements

- (68) kayen ši mġarba lli ka-y- eh y-řežŋ-u eh n-gul-u *revolutionair-en*
 EXIST INDEF Moroccan-PL REL ASP-3- er 3-become er 1-say-PL revolutionary-PL

deḍḍ t-taqafa dyał-hŋm
 against DEF-culture of-3PL

“There are some Moroccans who become revolutionaries against their culture.”
 (Samir)

- (69) ila ma qri-ti-š ma t-dir-š *oriëntatie* l had ki smit-u
 if NEG study-2SG-NEG NEG 2-do-NEG orientation to DEM how name-3M
 “If you haven’t studied, you’ve got no orientation toward this what’s it called.”
 (Hocine)

- (70) ka-y-dir waḥed *-opleiding* ŋla ši ḥaḥa
 ASP-3-do INDEF training about INDEF thing
 “He does some education about something.” (Najib)

- (71) waš kayen ši *verklaring* l dak š-ši
 Q EXIST INDEF explanation for DEM DEF-thing
 “Is there an explanation for this?” (Najib)

See also Samir's *waḥed l-verhaal ũla* "a story about" in (23) above. In the next two examples, the modification with *dyal* is considered a complement rather than a possessive construction, because the relationship between the noun and the following PP is analogous to that between a transitive verb and its Direct Object (for instance, in *the translation of a subject*, the relationship between *translation* and *subject* is analogous to that between the verb and the Object in *she translated the subject*).

- (72) ka-ne-ſmel l[?]-vertaling dyal waḥed l-madda
 ASP-1-work [DEF-]translation of INDEF DEF-subject
 "I'm working on the translation of a subject." (Zineb)
- (73) ḥna ſend-na hadak beeld dyal ..
 1PL at-1PL DEM image of
 "We've got this image of .. [of the mother country]" (Younes)

Dutch prepositional complements

Only two examples were attested of Dutch PP complements of an embedded Dutch noun, both of them in the contributions of Samir. Notice the analogy between examples (71) above and (74) below; *daarvoor* in (74) is a special form that occurs when a pronoun referring to an inanimate entity is the Object of a preposition (more details about this type of Dutch PP are presented in Chapter 6 section 2.6).

- (74) kayn-in daar-voor verklaring-en
 EXIST-PL there-for explanation-PL
 "There are explanations for that." (Samir)
- (75) ḥna ſend-na eigenlijk voorsprong op ander-en
 1PL at-1PL really advantage on other-PL
 "We really have an advantage on others." (Samir)

MA clausal complements

Various embedded nouns subcategorize for a MA complement clause introduced by the complementizer *baš*, followed by the subjunctive mood of the verb (marked by the absence of the indicate prefix *ka-/ta-*). Most examples involve nouns that express the modal concept of possibility. See also *gevoel baš* in (37) above, *gewoonte baš* in (39), and *het vermogen baš* in (122) in section 4.1 below.

- (76) ſend-ek zeſma kans, mogelijkhed-en, baš t-dir temmak ši eh ..
 at-2SG EPIST chance possibility-PL COMP 2-do there INDEF er

 t-dir ši t-dir toekomst creër-en
 2-do INDEF 2-do future create-INF

“You would have a chance, possibilities to er ..to er to create a future there”
(Younes)

- (77) šuf-i *capaciteit-en* lli ʔend-ek baš, baš t-dir ši ḥaža
look-IMP-F capacity-PL REL at-2SG COMP COMP 2-do INDEF thing
“Look what capacities you have to .. to do something!” (Younes)

- (78) ʔend-ek *kans* baš y-welli-w *meer intellectuel-en*
at-2SG chance COMP 3-become-PL more intellectual-PL
“There’s a chance that there will be more intellectuals.” (Najib)

- (79) ʔend-i *gevoel* baš ne-qra, walakin ..
at-1SG feel COMP 1-read but
“I have a feel for reading, but ..” (Hocine)

In this context, the MA complement clause marked by *baš* plus subjunctive is equivalent to Dutch complement clauses introduced by the complementizer *om* followed by *te* plus infinitive (e.g. *mogelijkheid om (..) te studeren* “possibility to study”). The same *baš* clause also occurs as a complement of Dutch embedded verbs (Chapter 6 section 4.2).

Dutch clausal complements

Only one incomplete example of an embedded Dutch noun with a Dutch complement clause was found in the data. In this unique example the speaker interrupted his flow after the complementizer.

- (80) kan-et ʔend-i *mogelijkhed-en om...*
be-3F at-1SG possibility-PL COMP
“I had possibilities to .. [study].” (Samir)

Thus, regarding complements of embedded Dutch nouns, whether prepositional or clausal, we can conclude that these tend to be in MA rather than in the language of the noun. Dutch complements - three in total - are only found in Samir’s contributions.

5.1.10 Attributive adjectives

Remarkably, hardly any embedded Dutch noun is modified by MA attributive adjectives. Two examples have been cited above as (9) and (52) and are (partly) reproduced here.

- (81) (xešš-ek f küll eh *wijk* wella küll derb y-kun-u sakn-in žuž, tlata,
must-2SG in every er district or every alley 3-be-PL live-PL two three

baraka!) la, huma ka-y-šift-u miyat turki u meğrabi f *wijk* waḥd-a!
 that's-all no 3PL ASP-3-send-PL hundred Turk and Moroccan in district one-F
 “In every district or every alley there should be two or three [immigrants], no
 more. No, they send a hundred Turks and Moroccans to one district!” (Hayat)

- (82) ḥna-ya mxellṭ-in temmak, *bawtlans* mxellṭ-in
 1PL-EMPH mixed-PL there foreigners mixed-PL
 “We are mixed there, mixed foreigners.” (Mustafa)

Both are atypical examples for different reasons. *waḥda* in (81) is formally an adjective but it is at the same time also a numeral. In (82) the embedded noun [bawtlans] is not like the other embedded nouns discussed in this chapter. It is a strongly assimilated form that became part of the speaker's MA before he learned to speak Dutch (in fact, Mustafa's mastery of Dutch was still defective at the time of the recordings).¹⁴ It is uncertain whether the speaker perceives the final -s as a plural marker. The word is probably based on the Dutch adjective *buitenlands* [bœytəland̥s] “foreign”, the Dutch word for “foreigner” being *buitenlander* [bœytəlandər] with plural *buitenlanders*. I found only one further instance, in Younes' data:

- (83) b š-šəḥḥ huwa *racist*, *racist* kbir hadak
 with DEF-reality3M racist racist big DEM
 “But he is a racist, a big racist he is.” (Younes)

Interestingly, in each of these three instances the Dutch noun modified by a MA adjective is repeated from the immediately preceding context. Later on it will become clear that repetition is often involved in atypical insertion types, see section 4 in Chapter 11.

The dearth of MA attributive adjectives modifying embedded Dutch nouns is striking. It contrasts with the unrestricted modification of embedded nouns by means of possessive constructions, adjunct or complement PPs, and relative clauses. Even if we consider that attributive adjectives may not be so frequent in informal spoken MA, they are at least more frequent than relative clauses.¹⁵ Indeed, one gets the impression

¹⁴ The word for “foreigner” is one of the first words to be acquired by immigrants, as it denotes their social group and identity in the context of the Dutch society.

¹⁵ In order to get an impression of the frequency of attributive adjectives in MA conversations I investigated randomly chosen monolingual MA stretches uttered by Mustafa and Warda, who represent the unschooled and highly educated varieties, respectively. For either speaker, I counted the types of modification for the first 100 nouns (tokens) in the MA texts that could possibly be modified by an adjective, that is, excluding adverbially used nouns and body part nouns in reflexive

that speakers make use of other modification strategies in order to avoid attributive adjectives. Thus in (1) at the beginning of this chapter, partially reproduced here as (84), Warda uses a relative clause *hiya ṣġira* to express “small” as an attribute of “village”.

- (84) ḥna sakn-in f dorp, hiya ṣġir-a
 1PL live-PART-PL in village 3F small-F
 “we live in a village, which is small” (Warda)

Likewise in (34), the relevant part of which is reproduced as (85), Samir makes use of a roundabout genitive construction to express “these Dutch workers”.

- (85) had l-*arbeider-s* dyal l-huḷandī-yin
 DEM DEF-worker-PL of DEF-Dutch-PL
 “these Dutch workers” (Samir)

In addition, embedded Dutch nouns are commonly accompanied by Dutch adjectives. This is discussed in section 2 below.

5.1.11 Relative clauses

EL Dutch nouns can freely be modified by MA relative clauses. Examples are found with many of the respondents, as shown in the above (1), (4), (13), (16), (22), (40), (55) and (77). Dutch relative clauses modifying embedded Dutch nouns are attested only in the case of Samir, cf. (35) and (47). More facts about relative clauses in the MA/Dutch corpus are presented in Chapter 8, (section 1.1) and Chapter 9 (section 3.1.1).

5.1.12 Interrogative forms

Samir frequently uses the interrogative construction [*nta ʃaš* N] for “what N; what kind of N”. This construction is also applied to Dutch nouns. Jamal provides one occurrence of a parallel construction with the alternative genitive particle *dyal*.¹⁶

pronouns, for example. The outcomes are as follows: of Mustafa’s 100 nouns, 6 were modified by an attributive adjective, 8 by possessive PPs, 1 by a complement PP and 2 by relative clauses. As for Warda’s 100 nouns, the numbers are: 8 attr. adj., 11 poss. PPs, 3 complement. PPs, and 2 rel. clauses. In the same stretches, Mustafa used 14 predicative adjectives and Warda 2.

¹⁶ The genitive particles *nta ʃaš* and *dyal* are dialect variants. Remarkably, Samir uses *dyal* in all contexts other than the interrogative forms discussed here. His frequent use of question words relates, of course, to his role as an interviewer in most of the recorded conversations.

- (86) ntaʃ aʃ *serie-s* ka-t-ʃuf-i ?
 of what serial-PL ASP-2-watch-F
 “What TV serials do you watch?” (Samir)
- (87) ntaya ntaʃ aʃ *vakk-en* ka-ye-ʃžeb-l-ek durka ?
 2M of what subject-PL ASP-3-please-to-2SG now
 “What [school] subjects do you like now?” (Samir)
- (88) dyal aʃ *voordeel* dyal-ek? *woordenschat* wella *accent* wella ..
 of what advantage of-2SG vocabulary or accent or
 [on the advantage of a Moroccan background for the study of Arabic]
 “What sort of advantage do you have? Vocabulary or accent or ..” (Jamal)

The construction [*ntaʃ* ~ *dyal aʃ* N] employed by Samir and Jamal seems to be a dialectal peculiarity. It serves the same function as the more common MA Koine forms [*ina* ~ *fina* ~ *ana* ~ *fana* N] and [*aʃ* ~ *waʃ men* N] (cf. Caubet, 1993, I: 170, 172). Yet, while the latter constructions can only be used attributively, *ntaʃ aʃ* and *dyal aʃ* may occur as predicates, as in the above example (88) and in (141) below.

5.2 Adjective-Noun combinations

Besides nouns and plural nouns, Dutch collocations of an adjective and a noun are regularly embedded. These have an internal Dutch structure which is apparent from the relative order of adjective and noun (Adj-N, as opposed to MA N-Adj), and agreement marking on the adjective. As a general rule, the gender agreement suffix on attributive adjectives consists of a final schwa *-e* [-ə] for all forms apart from the indefinite singular neuter. In addition, *-e* is omitted in certain idiomatic adjective-noun collocations and in the case of attributively used past participles that end in *-en* and certain irregular adjectives.

- (89) ma te-xdem, ġadi te-mʃi — *social-e dienst*
 NEG 2-work FUT 2-go [to] social-AGR service
 “If you don’t work, you’ll go [to] the Social Service.” (Younes)
- (90) ġadi n-gul-u t-kun-i mezyan-a temmak b hadik l-xedma,
 FUT 1-say-PL 2-be-F good-F there with DEM DEF-work
vast-e aanstelling, waʃ ġadi te-ržʃ-i wella la?
 permanent-AGR appointment Q FUT 2-return-F or NEG
 Let’s say you’ll be all right there with this job, a permanent appointment, will you return or not?” (Samir)

- (91) ġa-ne-mši n-dir *chemisch-e technologie*
 FUT-1-go 1-do chemical-AGR technology
 “I’ll go there and do chemical technology.” (Mimoun)
- (92) walakin daba d-derri ġadi ye-dxel *islamitisch-e school*, škun ġadi
 but now DEF-child FUT 3M-enter Islamic-AGR school who FUT
 ye-lqa?
 3M-meet
 “But now the child will go to an Islamic school, whom will he meet?”
 (Maryam)
- (93) xeşş-ni ma n-dir-š *sterk-e band ofzo*
 must-1SG NEG 1-do-NEG strong-AGR tie or-something
 “I should not create a strong bond or something.” (Abdellah)

Most examples are fixed and often idiomatic expressions; yet in other cases, the idiomatic character of the adjective-noun combinations is not so evident.

- (94) muhimm ka-te-kber u ka-t-fekker f eh .. *ander-e interesse-s*
 anyway ASP-2-grow up and ASP-2-thing in er other-AGR interest-PL
 “Anyway, you grow up and you think of other interests.” (Abdellah)
- (95) Šend-ek *groen-e boek* nta-ya, *groen-e boek* dya l-?islam, *eerst-e jaar?*
 at-2SG green-AGR book 2M-EMPH green-AGR book of DEF-Islam first-AGR year
 “Do you have the green book, the green book on Islam, first year?”
 (Samir)

On the other hand, the class of fixed expressions cannot be easily demarcated. This is especially apparent if we consider that an expression can be idiosyncratic to such an extent that it is common knowledge only to the speaker and the addressee. In (95) for instance, *groene boek* “green book” refers to a particular book with which Samir and his interlocutors, Najib and Hocine, are familiar. It is possible that within the ‘speech community’ of students following a particular course in Islamic Studies, *groene boek* was an idiomatic expression referring to this particular book.

Omission of the agreement suffix

The respondents who are less fluent in Dutch tend to omit the agreement suffix in embedded adjective-noun combinations. (96) below is Zineb’s response to Samir’s question in (90); note how the agreement suffix is lacking in Zineb’s rendering of the same expression *vaste aanstelling* “permanent appointment”. Likewise, compare *sociale dienst* as produced by Younes in (89) to Fatima’s *sociaal-__ dienst* in (97) below. Two further examples of missing agreement suffixes are given in (98) and (99).

- (96) ma kayen-š, ma ʃend-na vast-__ *aanstelling* f l-meḡrib!
 NEG EXIST-NEG NEG at-1PL permanent[-AGR] appointment in DEF-Morocco
 “There is no such thing, we don’t have permanent appointments in Morocco!”
 (Zineb)
- (97) te-mši l l-bank wella ši ḥaža wella *sociaal-*__ *dienst* wella (..)
 2-go to DEF-bank or INDEF thing or social[-AGR] service or
 “When you go to the bank or something, or to the Social Service or (..)”
 (Fatima)
- (98) l-ʃam l-lewwel hadik f *lager-*__ *school*
 DEF-year DEF-first DEM in primary[-AGR] school
 “The first year is in primary school.” (Mustafa)
- (99) ḥna ʃend-na *sociaal-*__ *controle*
 1PL at-1PL social[-AGR] control
 “We have social control.” (Hocine)

Apart from the peculiarities with respect to the agreement marker, embedded adjective-noun combinations are exactly similar to the inserted nouns discussed in section 1.

5.3 Insertion of adjectives

The distinction between attributive and predicative adjectives is central to the description. The insertion of attributive adjectives as modifiers of matrix language nouns is highly limited. In contrast there are no constraints on the insertion of adjectives as predicates in copula constructions; for that matter, the insertion of any kind of predicate in copula constructions has few restrictions (see section 4.5 below and Ch. 7 section 5).

5.3.1 Attributive adjectives

The corpus contains only two clear examples of embedded Dutch attributive adjectives. In addition, there is one instance where the matrix language cannot be ascertained, and one occurrence of the quantifier *weinig* “a little” which shares some properties with adjectives. All four are reproduced here.

- (100) kayn-in ši mḡarba, n-ʃerf-u-hūm, y-lebs-u triku *paars*,
 EXIST-PL INDEF Moroccan·PL 1-know-PL-3PL 3-wear-PL sweater purple

 xuxi, u hadak š-ši
 peach·coloured and DEM DEF-INDEF

“There are Moroccans - we know them - who wear purple and peach sweaters and all that stuff.”(Younes)

- (101) n-nas *intellectuel-e* xεṣṣ-hūm ye-t-kūwnn-u
 DEF-people intellectual-AGR must-3PL 3-MP-create-PL
 “The intellectual people have to be formed.” (Najib)

- (102) walakin *weinig* muhiba ka-teṣṭi-l-ek *meer eh vaardigheid*
 but little talent ASP-3F-give-to-2SG more er skill
 “But a little talent gives you more skill.” (Hocine)

In the next example by Samir it is impossible to determine whether *geestelijke xūbz* “spiritual bread” is really a MA or a Dutch constituent.

- (103) hada maši bḥal l-xūbz, b ṣ-ṣeḥḥ *geestelijk-e* xūbz, maši eh ..
 DEM NEG like DEF-bread with DEF-reality spiritual-AGR bread NEG er
 “This is not like bread, spiritual bread alright, but it is not er ..[essential]”
 (Samir)

In (100) and (101) the adjective follows the noun it modifies, in accordance with MA grammar; in (103) and (102), the adjective and the quantifier are pre-nominal, as in Dutch. Note that quantifiers in MA also precede the noun, so only (103) violates ML word order if we decide that MA is the ML. In (101) and (103) we find the adjectival agreement suffix *-e*, which is grammatical in Dutch for the plural noun in (101), and for the indefinite noun in (103) if it is of common gender (the translation equivalent of *xūbz* “bread” happens to be neuter in Dutch, so “spiritual bread” translates as *geestelijk brood*). The adjective *paars* in (100) does not have an agreement suffix; the Dutch translation of this constituent would be *een paars-e trui* “a purple sweater”. As for the quantifier *weinig* in (102), it is not allocated the agreement suffix in Dutch unless it is preceded by a definite article.

Since *geestelijke xūbz* in (103) differs from *triku paars* in (100) with respect to word order as well as agreement marking, I am inclined to view the former as a Dutch NP with an inserted MA noun. Insertion of MA words other than culturally specific terminology is uncommon, but the repetition of a word as is the case with *xūbz* in (103) happens to be a factor that leads to the embedding of MA nouns in Dutch constituents (cf. section 1.2 in Chapter 9).

Regarding the agreement suffix *-e* in Najib’s example (101), I conjecture that it results from a tendency amongst non-fluent speakers of Dutch like Najib to generalise the suffixed allomorph of the Dutch adjective to all contexts. Consequently, in similar learners’ varieties of Dutch the adjective ending in the schwa vowel is actually an

unvariable and hence uninflected form.¹⁷ Another possibility is to consider [Intɛlɛktywɛlθ] as the homophonous plural noun *intellectuelen* “intellectuals” (the final /n/ is never pronounced) occurring in apposition to the MA NP *n-nas* “the people”.

5.3.2 Predicative adjectives

Embedded predicative adjectives such as *links* “leftist” in (46) above occur far more frequently in the data. MA predicative adjectives like attributive adjectives agree in gender and number with the antecedent on which the predication is made (usually the Subject, except in expressions like “to consider x as y”, see ex. (109)). In Dutch, the uninflected form of the adjective is used for both genders and numbers. This form is also used in the context of MA/Dutch CS.

A single adjective is a possible (predicate) constituent in Dutch. Moreover, embedded Dutch predicative adjectives are usually modified by Dutch degree adverbs, if any, see (104), (105). For these reasons I consider the insertion of predicative adjectives as (a subtype of) the insertion of predicate constituents, the latter being a relatively unconstrained insertion type; see also section 4.5 below and Ch. 7, section 1.4.

- (104) ɾa-h y-kun *heel lief, heel aardig*
 PRES-3M 3-be very gentle very kind
 “He will be very gentle, very nice.” (Fatima)

- (105) mma mma *ja n-gul-u eh veel te progressief*
 mother mother yes 1-say-PL er much to progressive
 “My mother, my mother is, well, let’s say, far too progressive.” (Samir)

(106) is the only counter-example in the data, where a Dutch adjective is modified by a MA degree adverb. Notice that *oud* “old” in this example is a repetition in Dutch of the MA term *kbir* “old”.

- (106) n-kun kbir, n-kun gaŋ kbir, šwiya *oud, oud bezzaf*
 1-be old 1-be completely old a-little old old very
 “I’ll be old, very old, a little old, very old.” (Abdelkrim)

¹⁷ The reality is more complex: on the one hand speakers like Najib generalise the inflected form using it even in predicative adjectives where it never occurs in native speakers’ Dutch; on the other hand they do not consistently use the suffixed forms even in appropriate contexts. This aspect of their Dutch grammatical system is in fact variable and probably subject to constant change.

Subcategorisation

Some embedded Dutch adjectives subcategorize for a prepositional complement:

- (107) ka-t-kun *moeilijk* ʃla ši wahed
 ASP-3F-be difficult for INDEF one
 “It is difficult for someone.” (Fatima)
- (108) kull-ši mužud walak inta-ye-bqa l-ʔinsan eh *niet tevreden*
 everything present but ASP-3-remain DEF-human-being er not satisfied

 matalen b l-.. eh
 for-example with DEF- er
 “Everything is there, but one remains dissatisfied for instance with er ..”
 (Najib)
- (109) ka-t-šuf-ha *heel anders* ʃla hna
 ASP-2-see-3F very different from here
 “You see that it is very different from here.” (Zineb)
- (110) nta mesʔul dyal-hūm, nta *verantwoordelijk* ʃla ..
 2M responsible of-3PL 2M responsible for
 “You are responsible for them, you are responsible for ...” (Najib)
- (111) d-dariža hiya n-gul-u *afgeleid* dyal eh .. l-ʃerbiya
 DEF-Moroccan-Arabic 3F 1-say-PL derived of er DEF-Arabic
 “Moroccan Arabic is let’s say derived from er Classical Arabic.” (Samir)

These cases are analysed as embedded Dutch Predicate constituents in which a MA PP constituent is inserted. In (107), (108) and (109) the complement prepositions *ʃla* and *b* are what one would expect in the MA translation equivalents. In (110) the preposition *ʃla* follows as expected after *mes ʔul* “responsible”, while *mes ʔul dyal* appears anomalous. Then in (111) the preposition *dyal* suggests calque-like influence of the Dutch expression *afgeleid van* “derived from”. In many contexts, particularly in genitives, the Dutch preposition *van* can be translated as MA *dyal*, although to mark the origin of something as in the above example one would normally use the MA preposition *men*. In (112) below, finally, the complement is in Dutch.

- (112) ila kan-u mehlul-in baš ye-tʃellm-u ši haža men huḷanda kun
 if be-3PL open-PL COMP 3-learn-PL INDEF thing from Holland APODOSIS

 kan-u daba *gelijk met die mensen in marokko*
 be-3PL now equal with those people in Morocco
 “If they had been open to learn something from Holland, they would now have come as far as those people in Morocco.” (Maryam)

Comparative forms

There are a few occurrences of embedded comparative adjectives as predicates. These are interesting because MA and Dutch have quite different syntactic constructions for comparisons. Dutch employs inflectional comparative and superlative forms for adjectives while analytic forms with the adverbs *meer* and *meest* “more, most” also occur. The second member (complement) of the comparison is introduced by one of the particles *dan* or *als*. (The Dutch word *anders* “different” in (109) has the same complementation pattern.) In MA only a small set of adjectives has inflectional comparative forms; the complement is marked by the preposition *men*. For the other MA adjectives, the positive form is used followed by the preposition *la* to introduce the complement. In addition, analytic forms with *kter* “more” occur. Superlative forms are marked by the definite prefix *l-*.

Firstly in (113) we find a Dutch complement of the comparative *slimmer* “more intelligent” introduced by the particle *dan*.

- (113) m̩m-u hiya ʃemmer-ha ma q̩ra-t u hiya eh hiya *slimm-er*
 mother-3M 3F ever-3F NEG study-3F and 3F er 3F intelligent-COMPAR

dan z'n vader

than his father

“His mother never studied and yet she er .. She is more intelligent than his father.” (Samir)

Now consider *beter* “better” in (114). This word has a complement clause introduced by the particle *ma* followed by the subjunctive verb (without the aspectual prefix *ka-*). This particular kind of complementation is analogous with the MA words for “better” *hsen* and *xir*.¹⁸

- (114) yeʃni ka-t-kun *afgestudeerd*, t-kun *werkloos*, *beter* ma t-kun
 that-is ASP-2-be graduated ASP-2-be unemployed better PARTICLE ASP-2-be

werkloos u ma der-ti-š afstuder-en

unemployed and NEG do-2SG-NEG graduate-INF

“So it is better to be graduated and unemployed than to be unemployed without being graduated.” (Hocine)

(115) is also an interesting example. “as ... as” is expressed in Dutch by the adjective and two particles: [*zo* Adj *als*]; in MA the positive form is used with the preposition *bhal* “like” to mark the second member.

¹⁸ Cf. the entries *hsen* and *xir* in Colin (1993, vol.2, 3): *nmut, hsen ma n ʃiʃ medlul* “I would rather die than live humiliated”.

- (115) f l-meğrib maši zo *moeilijk* bħal hna-ya
 in DEF-Morocco NEG as difficult like here-EMPH
 ‘‘In Morocco it’s not as difficult as it is here.’’ (Fatima)

Finally, consider the superlative form *de heiligst* in (140) hereafter. This form, *de heiligst* DEF·C/PL holy-SUPERLATIVE ‘‘the holiest’’ is not in accordance with Dutch grammar. In Dutch the superlative adjective is preceded by the neuter definite article when used predicatively: *het heiligst*. Alternatively the attributive form is possible, preceded by the plural definite article and inflected with the final schwa for agreement, with the optional elision of the redundant noun: *de heiligst-e (dagen)* DEF·PL holy-SUPERLATIVE-AGR (day-PL) ‘‘the holiest (days)’’.

5.4 Insertion of Nominal Constituents

An inserted NP is a full constituent in Dutch that has the distribution of an NP in the MA clause. What counts as an NP in Dutch and in MA is determined by distributional properties. There are different types of NP constituents which do not all have the same distribution. According to the above definition a single noun can be considered an embedded NP in certain contexts: both Dutch and MA allow for single noun predicates in copula construction, for instance. In this description of MA/Dutch CS, nominal constituents of this type are subsumed under the heading of ‘noun insertion’ (section 1). Analogously, embedded NPs that consist of just a noun and an attributive adjective are not distinguished from the insertion of ‘adjective- noun combinations’ (section 2). Here we will be concerned with types of embedded Dutch NPs that cannot be analysed as instances of content morpheme insertion. These are far less common than embedded nouns.

The description is ordered according to two criteria: Firstly I distinguish between embedded NPs in copula constructions, to be discussed at the end in 4.5 above, and NPs in other contexts. The NPs in other contexts are ordered according to their internal structure. The typical ‘complex’ Dutch NP consists of (at least) a determiner and a noun. With respect to the insertion of constituents the most interesting question arising is whether inserted constituents contain elements, particularly function words, that do not correspond to a similar category in the matrix language at the clause level. In order to investigate this embedded NPs are divided according to the type of determiner. Three more or less recurrent types of embedded Dutch NPs emerge from the data: nouns determined by a definite or indefinite article; nouns determined by possessive pronouns; and nouns determined by a quantifier, in particular, by a numeral. In addition, embedded pronouns are also classified as NPs by virtue of their distributional properties.

5.4.1 Nouns determined by a definite or indefinite article

Apart from copula constructions I noted 20 NPs containing a Dutch article. Nine instances were produced by Hocine, seven by Samir, while Abdellah, Maryam, Hayat and Mimoun contributed one instance each. This means that the insertion of this type of NPs is remarkably frequent in Hocine's utterances, while for the other respondents the frequency is on average less than once per conversation. Note that, since Samir participated in nearly all conversations, seven instances by Samir actually amount to the same frequency as one instance by Abdellah, for instance. In the following examples the Dutch articles are glossed as INDEF "indefinite (singular)", DEF·C "definite common gender", DEF·N "definite neuter" and DEF·PL "definite plural".

- (116) ġadi ne-fhem *de* *hoofdlijn-en*, zeg maar, *die* begrijp ik wel
 FUT 1-understand DEF·PL outline-PL say just those understand I AFFIRM
 "I'll understand the outlines, that is, I do understand those." (Abdellah)

- (117) daba ka-ne-qra *een roman* Ħend-i *de roman* Ħta-ha-ni *de buurman*
 now ASP-1-read INDEF novel at-1SG DEF·C novel give- 3F-1SG DEF·C neighbour
 "Now I'm reading a novel. I have a novel that my neighbour gave me."
 (Hocine)

- (118) ka-t-kun f waħed eh n-gul-u eh *een soort ruimte*, zeg maar, waar
 ASP-2-be in INDEF er 1-say-PL er INDEF kind space say just where

eigenlijk niks is
 actually nothing is

"You'll be in a er .. let's say er a kind of space, that is, where there's actually nothing." (Samir)

- (119) ila huḷandi, ila gal ana ma bġi-t-š weld-i, bent-i te-t-zewwež
 if Dutchman if say 1SG NEG want-1SG-NEG son-1SG daughter-1SG3F-MP-marry

t-žib-l-i ši *surinamer of een zwart-e neger*, (dan hebben
 3F-bring-to-1SG INDEF Surinamese or INDEF black-AGR Negro then have

— *ook gelijk, eigenlijk*

[they] also right actually

"If a Dutchman, if he says I don't want my son, my daughter to marry, to bring home some Surinamese or a black Negro, then [they] are right too, really."
 (Hayat)

Nine examples are inherently definite nouns such as "the truth", "the reality", "the content" or "the first / second / next / last / only one", cf. (120)-(122). (Concerning the subcategorisation of *vermogen* "capacity" in (122), see section 1.9 on noun complements.)

- (120) *ila bġi-tu t-reyyh-u ʃend-i, t-šerb-u l-atay u t-neʃs-u u te-mši-w*
 if want-2PL 2-rest-PL at-1SG 2-drink-PL DEF-tea and 2-sleep-PL and 2-go-PL

eh de volgend-e dag t-zid-u!

er the next-AGR day 2-go-on-PL

“If you (PL) want to stay at my place, you drink tea, you sleep, and you leave
 er the next day you go on.” (Samir)

- (121) *ka-y-šuf eh schaduw dyal de werkelijkheid*

ASP-3-see er shadow of DEF-C truth

“He sees the shadow of the truth.” (Samir)

- (122) *ʃend-i het vermogen baš n.. eh ne-tbeʃ chimie*

at-1SG DEF-N capacity COMP 1- er 1-follow chemistry(Fr)

“I have the capacity to study chemistry.” (Mimoun)

In (117) above, the definite article in *de roman* “the novel” seems anomalous in both languages in this context: the particular novel that Hocine got from his neighbour was not mentioned before in the conversation, and I suspect that Hocine’s addressee does not know the referent of this NP *de roman*.¹⁹ In the following example, also produced by Hocine, the use of a definite article is in accordance with MA rather than Dutch grammar.

- (123) *ka-thess b l-luġa, ʃend-ek het gevoel*

ASP-2-feel with DEF-language at-2SG DEF-N feel

“You feel the language, you have a feel [for it].” (Hocine)

5.4.2 Nouns determined by a possessive pronoun

I found five Dutch NP of this type in MA clauses, two in Samir’s contributions, and the other two produced by Hocine and Fatima. Four instances are reproduced here, also see *jouw vergoeding* “your reward” in (59) above.

- (124) *gal-l-ha: “ana ġadi ne-ddi-k l jouw kamer” u mša nʃes mʃa-ha*
 say-to-3F 1SG FUT 1-take-2SG to your room and go sleep with-3F

“He said to her: ‘I’m going to bring you to your room’, and he went [there]
 and slept with her.” (Hocine)

¹⁹ Besides, after a definite NP the relative clause marker *lli* “is mandatory if the subordinate clause is to be interpreted as a restrictive adjectival modifier of the prestated topic” (Harrell, 1962: 165). So if the definite article is intended, then the omission of *lli* is anomalous.

- (125) š-šurut dya*l*-i huwa ʔanna-ni t-kun ʃend-i *mijn vrijheid*
 DEF-condition·PL of-1SG 3M COMP-1SG 3F-be at-1SG my freedom
 “My condition is that I should have my freedom.” (Fatima)
- (126) kull wa*h*ed ʃend-u, ʃend-u *zijn interesse-s*
 every one at-3M at-3M his interest-PL
 “Everyone has, has his own interests.” (Samir)

The last example shows a particular type of possessive with the word *eigen* “own”, which refers to a co-referent in the sentence, in this case the Subject. MA has no distinct construction to emphasise the possessor.

- (127) ma ka-te-qra-y-š ze*ʃ*ma baš t-kun ʃend-ek *eigen inkomen, eigen*
 NEG ASP-2-study-F-NEG EPIST COMP 3F-be at-2SG own income own
salaris, eigen flus?
 salary own money·PL
 “Is it not so that you study in order to have your own income, your own salary, your own money?” (Samir)

5.4.3 Nouns determined by a quantifier

Here we distinguish between numerals and other quantifiers. With the exception of the assimilated Dutch word *xulda* “guilder”, the respondents show a preference for counting Dutch nouns with Dutch numerals: 8 tokens of embedded Dutch nouns plus MA numeral (see section 1.6 above), as against 9 tokens of Dutch nouns plus Dutch numeral. This makes counted nouns a relatively frequent type of inserted nominal constituent. In addition, there are seven instances of Dutch nouns with other quantifiers (“only, many, more, most, all”). In this category of embedded NPs, Hocine again contributes a large amount of the tokens (4 numerals, 4 other quantifiers).

Examples with numerals (see also (136) hereafter and (7) in Chapter 7):

- (128) dar-u-k *één cultuur*
 do-PL-2SG one culture
 “They placed you [in] / made you one culture.” (Najib)
- (129) ana šex*ʃ*iye*n* ma ġadi-š ne-qbel matalen huwa ye-m*ʃ*i b *tien meter*
 1SG personally NEG FUT-NEG 1-accept for·example 3M 3-walk with ten metre

u ana ʕada muṛa-h žayy-a

and 1SG finally behind-3M come·PARTICIPLE-F

“I personally will not accept that he walks ten metres [ahead of me], while I’m coming behind him.” (Zineb)

- (130) xeṣṣ-na n-xerž-u men hnaya, __ *tien minuut-je-s*
 must-1PL 1-go·out-PL from here [in] ten minute-DIM-PL
 “We have to get out of here [in] ten minutes.” (Jamal)

Examples with other quantifiers (see also *meer vaardigheid* in Hocine’s statement (102)):

- (131) *de meeste schrijver-s, arabisch-e schrijver-s, dda-w d-duktura*
 DEF·PL most writer-PL Arab-AGR writer-PL take-PL DEF-doctorate

dyal-hūm f sorbonne

of-3PL in Sorbonne

“Most writers, Arab writers, got their doctorate at the Sorbonne.”

(Hocine)

- (132) ġir r-ražel ʕend-u *meer mogelijkheid-en dan de vrouw-en, ook hier*
 only DEF-man at-3M more opportunity-PL than DEF·PL woman-PL also here

in Nederland

in the·Netherlands

“Men just have more opportunities than women, even here in the Netherlands.”

(Fatima)

- (133) u nta, ta nta ʕend-ek *alleen maar buitenlander-s?*
 and 2M also 2M at-2SG only just foreigner-PL
 “And you, do you also have only foreigners [as friends] ?” (Samir)

5.4.4 Pronominals

Because of their distributional properties, pronouns are considered nominal constituents. Including those in copula constructions (see (139)), six inserted pronominal forms were found, which were not personal pronouns; *iedereen* “everyone” in (134) and (135), the independent demonstrative *dit* “this” in (136) and *mezelf* “myself” and *iemand anders* “someone else” in (139) below.

- (134) *iedereen* ʕžeb-u l-ḥal, ḥetta d-ḍar ʕžeb-hūm l-ḥal
 everyone please-3M DEF-situation even DEF-family please-3PL DEF-situation
 “Everyone liked it, even my family liked it.” (Fatima)

- (135) *ra-h iedereen ka-ye-kteb š-šeŕ*
 PRES-3M everyone ASP-3-write DEF-poetry
 “Everyone writes poetry.” (Fatima)
- (136) *u ġadi t-dir twee punt-je-s, dit en dit en dit, b l-ŕeḃiya*
 and FUT 2-do two point-DIM-PL this and this and this with DEF-Arabic
 “And you will do two items, this and this and this, in Arabic.” (Hocine)

5.4.5 NPs in copula constructions

In general predicates are freely inserted in copula constructions. Apart from predicates that consist of single nouns or adjective-noun combinations (see (4), (27), (28), (67), (68) and *intellectuelen* in (78)), there are fifteen embedded NP predicates in the data corpus, most of them produced by Samir. The most common MA copula is the verb *kan* “to be”. When no marking for mood, modality, aspect, or tense is required, this verb is replaced by either a ‘zero copula’, for indefinite predicates, or a third person pronoun in the case of a definite predicate (cf. (140)). Consider the following examples of embedded NP predicates.

- (137) *hadak eh volgens mij echt een racist*
 DEM er according.to me really a racist
 “According to me, that one is really a racist.” (Younes)
- (138) *kūn-t heel streng moslim*
 be-1SG very strict Muslim
 “I was a very strict Muslim.” (Samir)

In (137) the predicate NP consists of an article plus noun. In Dutch some predicate nouns that designate functions, offices, professions, ideologies et cetera display adjective-like behaviour (Wetzer, 1995: 114 n.7). This accounts for the absence of an indefinite article in the NP *heel streng moslim* “a very strict Muslim” in (138). Note that *heel streng moslim* is a perfect predicate complement in Dutch but it constitutes no evidence for NP insertion as distinguished from the insertion of adjective-noun combinations, even though in this case the adjective is modified by the so-called limiting adverb *heel* “very”. In the passage reproduced below, we find the pronominal forms *mezelf* “myself” and *iemand anders* “someone else” as insertions.

- (139) *mŕa-k kūn-t mezelf, b š-ŕeḃḃ mŕa ḃḃa u mŕa mḃa u mŕa*
 with-2SG be-1SG myself with DEF-reality with father and with mother and with
 xu-ya - xu-ya waxxa - f l-mustaqbel lḷahu ŕalem, ġadi n-kun
 brother-1SG brother-1SG alright in DEF-future God know*part FUT 1-be

mezelf, b ṣ-ṣeḥḥ durka ṣawda ana-ya iemand anders, ya-k?
 myself with DEF-reality now still 1SG-EMPH someone different QTAG-2SG
 “With you I was being myself, but with my father and my mother, and with
 my brother - my brother ok - well, maybe in the future God knows, I will be
 myself. But now I am still being someone else.”

In two of Samir’s utterances, the NP on which the predication is made is an insertion. Notice the pronominal copula in (140). The embedded question in (141) is a copula construction, with the question ‘word’ *nta ṣ aš* (on which see 1.12 above) as the proposed predicate, and *de inhoud dyal l-hedra* as Subject.

(140) *de laatste tien dagen, hè, dyal remḍan huma de heilig-st*
 the last ten days QTAG of Ramadan 3PL the holiest-SUPERLATIVE
 “The last ten days, you know, of Ramadan, these are the holiest.” (Samir)

(141) *walakin ma ḡadi-š t-šuf nta ṣ aš de inhoud dyal l-hedra*
 but NEG FUT-NEG 3F-look of what the content of DEF-discourse
 “But she will not look into what the content of the discourse is like.”
 (Samir)

Both (140) and (141) are analysed as cases of layered embedding: *de laatste tien dagen dyal remḍan* and *de inhoud dyal l-hedra* are inserted Dutch NP constituents that contain an Arabic PP constituent in the form of the analytic genitives with *dyal*.

5.5 Grammatical functions of EL nouns and nominal constituents

There are no constraints in terms of grammatical functions on the distribution of embedded nouns as compared with MA (matrix language) nouns. In order to illustrate this, an inventory is presented here of the examples cited in this chapter, ordered according to their grammatical function. The cited adjective-noun combinations and embedded constituents are listed here as well. There is no apparent reason why the distribution of adjective-noun combinations would be different from that of the nouns.

The embedded nominal constituents, on the other hand, form a heterogeneous group, and it is likely that various NP types have a distribution that diverges from that of single nouns or adjective-noun combinations. (For instance, nouns determined by a quantifier but without a definite article are indefinite, and indefinite NPs are less frequent as Subjects.) Because of the heterogeneity of this insertion category and the small number of examples, I will not draw any conclusions with regard to the distribution of embedded NPs.

Unclear cases have been left out. When more than one noun or NP is embedded in the example at stake, the appropriate item is specified here; otherwise I only refer

to the number of the example. N stands for embedded nouns, Adj-N for adjective-noun combinations and NP for nominal constituents.

i. Subject

- a. subject of a verb: N (7), (21), (29); NP *de buurman* in (117), (131), (135)
- b. subject of *ſend* “to have”:²⁰ N *vrouw* in (6); *plato* in (23)
- c. subject of a copulative construction: N (3), (5), *voordeel* in (88); NP *de laatste tien dagen* in (140), (141)
- d. subject of an existential construction: N (2), (13), (71), (74)

ii. Predicate of a copula construction

N (4), (27), (28), (67), (68), *intellectuelen* in (78), (83); NP (137)-(139)

iii. Direct Object complement

- a. direct object complement of a verb: N (15), (18), *lijst* in (24), (30), (35), *opleiding* in (31), (36), (49), (50), five nouns in (54); *vergoeding* in (59); *vrijstelling* in (65); (66), (69), (70), (72), (77), (86), (87), *schaduw* in (121); Adj-N (91), (92), (93); NP (116), *een roman* in (117), *een zwarte neger* in (119), (136)
- b. object on *ſend* “to have”: N (11), *theorie* in (23), (26), (37), (39), (40), (47), (73), (75), (76), *kans* in (78); (79), (80); Adj-N *groene boek* in (95), (96), (99); NP *de roman* in (117), *het vermogen* in (122), (123), (125), (126), (127), (132), (133)

iv. Prepositional complement, where the PP is

- a. complement of a verb: N (25), (32), (38), (43), (46), (63); Adj-N (89), (94), (97); NP (124)
complement of a noun: N (64)
- b. an adjunct (locative or time): N (1), *straat* in (8), *grot* in (23); (41), (42), (44), (48), (52), (53); Adj-N: *vrije ruimte* in (65), (90), (98); NP (118), (129)
- c. predicate of a copula: N (16), (45); NP (137), (138), three pronouns in (139),
- d. an analytic genitive construction: N *leraar-opleiding* in (31); *koffie* in (50); (51), (56); NP *de werkelijkheid* in (121)

v. Possessor in a synthetic genitive construction (construct state): N (33)

vi. Adverb (time): N (20), (57), (58); NP (120)

²⁰ The construction *ſend* plus pronominal suffix “to have”, common to most varieties of Arabic, displays features of both verbs and prepositions, but is best characterised as a verb with the possessor as Subject and the possessed item as grammatical Object. Comrie discusses this matter in detail (Comrie, 1989: 219-24, and in a number of articles).

In addition, we note that embedded nouns occur as left-dislocated Topics, the referent of which has various grammatical functions in the ensuing clause:

vii. Left-dislocated Topic: **N** *knikker* in (8); (12), (19), (34), (55); **NP** (134)

This overview shows that there is no indication that embedded nouns are restricted to particular grammatical functions. Not included in this list are Indirect Objects, which are marked by *l* in MA. This is because this kind of complement is restricted to very few contexts, namely verbs that take two Objects, such as *ḥa* “to give” and *werra* “to show”.

However, we observe an idiosyncrasy with respect to the preposition *l* in other contexts; this will be discussed in the next section.

5.6 Omission of the preposition *l*

Just like the definite prefix *l-*, the MA prepositional clitic *l* that marks nominal constituents for the thematic roles of DIRECTION/GOAL and RECIPIENT/BENEFACTIVE is often omitted before inserted Dutch words. In the case of embedded Dutch nouns, omission of the preposition often, if not always, co-occurs with the omission of the MA definite prefix. The omission of the preposition *l* is found with several respondents, both after the verb *mša* “to go” and after other verbs, as the following examples illustrate (see also ex.(89)). The missing prepositions and definite prefixes are indicated between square brackets.

(142) ḡadi-n ne-mši-w _ _-*bibliotheek*
 FUT-PL 1-go-PL [to DEF-]library
 “We’ll go to the library.” (Hocine)

(143) mši-t _ _-*kleuterschool*
 go-1SG [to DEF-]nursery-school
 “I went to the nursery school.” (Hayat)

(144) mša _ _-*discotheek*
 go [to DEF-]disco
 “He went to the disco.” (Samir)

(145) ḡlaš huwa ka-y-dexxlu-h _ *VWO* [feweo] w-ana ma y-dexxlu-ni-š ?
 why 3M ASP-3-enter-PL-3M [to] VWO and-1SG NEG 3-enter-PL-1SG-NEG
 “Why do they let HIM go to the VWO [pre-university education], and not me?” (Samir)

- (146) maši ana, ɖəɾb-u-l-i tilifun ana, ɖəɾb-u __ __-directie temmak
 NEG 1SG hit-PL-to-1SG telephone 1SG hit-PL [to DEF-]management there
 “Not me, they didn’t call ME, they called the management there.”
 (Younes)

Remember that in certain contexts we cannot reliably determine whether it is the definite prefix, the preposition or both that surface, due to the difficulties concerning the perception of geminate consonants (cf. section 1.4 above on definiteness). Further note that only the definite prefix, but not the preposition, assimilates to the initial coronal consonant (so *d-ɖar* “the house”, *l d-ɖar* “to the house”). In (142) and (143) both the preposition and the definite prefix are clearly lacking; in (144) and (146) the preposition is clearly lacking, while the presence of the definite prefix depends on the perception of /d/ in *discotheek* and *directie* as a geminate cluster. Yet, if we conjecture that the omission of the clitic preposition *l* is due to it being unproductive before foreign words, we must assume that the definite prefixes are absent as well. Otherwise the preposition *l* would attach to the MA definite prefix.

The same respondents who sometimes omit the preposition *l*, do attach it to Dutch nouns on other occasions; compare the following two examples from Hocine and Samir with those above. Note, however, that *l* in *l arbeidsburo* in (148) may also represent the definite prefix.

- (147) nhaɾ l-ʒemʃa ɣadi-n ne-mši-w l amsterdam
 day DEF-Friday FUT-PL 1-go-PL to Amsterdam
 “Friday we’ll go to Amsterdam.” (Hocine)

- (148) ɣin mši-t l arbeidsburo ʃta-w-l-ek had š-ši eh ..
 when go-2SG to employment-office give-PL-to-2SG DEM DEF-INDEF er ..
 “When you went to the employment office, they gave you this er ..”
 (Samir)

The omission phenomenon is characteristic of *l*. This preposition is missing in a large number of the obligatory contexts, while the other unisegmental prepositions surface in the appropriate contexts with only few exceptions (see below). The preposition *l* further contrasts with *f* (and probably *b* and *d* as well), in that the string *fl*- [in DEF-] is realised as a phonological word [fəɫ] in hesitation contexts (cf. 1.5), unlike the string *ll*- [to DEF-]. Instead, the two morphemes in *ll*- merge into a single geminate consonant. It turns out that the preposition *l*, like the definite article, behaves like a semi-productive prefix.

As mentioned earlier, the preposition *f* (locative) was also found to be missing in some cases. However, this frequently occurring preposition was omitted only five times. Except for (149), the examples containing omissions involve embedded Adjective-Noun combinations (for which see section 2).

- (149) ne-mši-w _ *acht uur dreëntwintig*. řa-h mša hadak.
 1-go-PL [at] eight hour twenty-three PRES-3M go DEM
 “We’ll go at eight twenty three. That one’s gone.” (Jamal)

- (150) sir t-qelleb řla ři bent _ _-*zelfde niveau ongeveer* bař
 go•IMP 2- search for INDEF girl [at DEF-]same level about COMP

 t-fehm-u beřđiyat-kũm
 2-understand-PL RECIPROCAL-2PL
 “Go look for a girl of about the same level, so that you will understand each other.” (Abdellah)

- (151) xeřř-ek t-kun ta hiya zeřma _ _-*zelfde situatie*
 must-2SG 3F-be also 3F EPIST [in DEF-]same situation
 “She has to be in the same situation.” (Samir)

The string *ři bent zelfde niveau* in (150) can alternatively be interpreted as a synthetic possessive “a girl of the same level”, since *bent* is one of those nouns that may occur as the first element in this construction. In (150) and (151), a definite article is also lacking, unless the speakers attribute the same status to Dutch *zelfde* “same” as to MA *nefs l-*, which appears in the MA construction [*nefs l-N*] “the same X”.

5.7 Summary

Embedded Dutch nouns have been investigated in detail in order to determine to what extent they behave as matrix language counterparts. It becomes apparent that in some respects the Dutch nouns have a different pattern of use. Firstly, there is little evidence that MA morphological processes are productive with Dutch stems. Dutch nouns are not inflected for plural, Dutch plural nouns being used instead. The MA definite prefix *l-* is the only ML affix that regularly attaches to Dutch nouns, however this prefix is also left out in most contexts. The omission of *l-* is almost constant when it is redundant, viz. after the indefinite article *wahed* and after any of the demonstrative determiners. I also argued that the prefix *l-* is more likely to surface before Dutch nouns when *l-* constitutes a phonological word with a preceding determiner or preposition, although this only became apparent in the data of Samir. Secondly, ML and EL nouns differ in that the latter are hardly ever modified by an ML attributive adjective. Inversely, Dutch adjectives are seldom inserted in the attributive function, whereas there appears to be no restriction on the insertion of predicates. On the other hand, the insertion of Dutch nouns modified by a Dutch attributive adjective is common and widespread. These facts suggest that there are specific ties that bind nouns and attributive adjectives from the same language; I will return to this in Chapter 11. If we disregard the unproductive MA definite article and adjectival modification, then there is no indication that the distribution of Dutch nouns

in MA clauses is different from that of MA nouns in terms of grammatical functions or modification patterns. Dutch nouns are attested in all major syntactic positions in the MA clause.

Concerning embedded NP constituents, these turn out to occupy a rather marginal position in the respondents' CS repertoire. Of the various types of nominal constituents having the distribution of an NP in Dutch, only a few are regularly inserted in MA clauses. In particular, when embedded Dutch nouns are counted, Dutch numerals are as common as MA ones. This seems to be a weaker version of the tendency already observed for nouns and attributive adjectives to be in the same language. In addition, there were some instances of embedded indefinite pronouns, nouns modified by other Dutch quantifiers or possessive pronouns. In Samir's data Dutch nouns determined by a Dutch article occur as predicates in MA copula constructions. Apart from this, only one respondent, Hocine, recurrently inserted Dutch NPs containing Dutch articles. In fact, Hocine contributed a large amount of the embedded NPs in general. The insertion of predicate NPs resembles the unconstrained embedding of adjectival predicates. These findings show that copula constructions should be carefully distinguished from matrix clauses with other verbs: the insertion of predicates in a copula construction is fairly unrestricted, whereas the insertion of NPs in other positions is subject to many constraints.

The MA preposition *l* cliticizes to nominal constituents to mark Indirect Objects and complements of motion verbs, that is, syntactic functions that usually coincide with such semantic roles as GOAL or BENEFICIARY. It turns out that this preposition often fails to surface before Dutch words. In this respect the preposition *l* behaves similarly to the definite prefix *l-*.

Evaluating the findings in the light of the MSA, we note the following points. The absence of certain MA function morphemes, namely the definiteness marker *l-* and the preposition *l*, results in so-called bare forms. As I acknowledged in Chapter 2 (section 4.1.1), bare forms constitute a recurrent feature of CS varieties that cannot be accounted for by the notions of matrix structure and embedded categories. The unproductiveness of certain matrix language morphological processes calls for an independent explanation.

The observation that Dutch attributive adjectives and most types of NPs are hardly ever inserted does not constitute a serious threat to the MSA. For the time being, the non-insertion of certain categories of EL material can be ascribed to lack of congruence between EL and ML categories, although this obviously raises the question of what congruence really is.

As stated above, the data point to a special dependency relationship between attributive adjectives and the nouns that they modify. It appears that the selection of attributive adjectives is not as unrestricted as the selection of nouns, or indeed largely depends on the noun being modified. This dependency relationship is somewhat similar to that between verbs and the prepositions for which they subcategorise (cf. Chapter 6 sections 3.4 and 3.5). Dependency relationships complicate the image of CS as the insertion of congruent categories, but are not at odds with it. Then the EL

word order in embedded adjective-noun combinations reveals that embedded content words can import certain EL grammatical features into the ML structure. The complementation patterns of certain nouns and predicative adjectives also demonstrate this. In sum, there is unequivocal evidence that insertion is not restricted to phonological forms but extends to EL grammar as well, although the amount of EL grammatical features accompanying embedded forms may vary from one instance to another.

Chapter 6

Dutch Verbs and their Complements in MA

Dutch verbs are integrated in MA clauses by means of a so-called periphrastic *do*-construction with the MA verb *dar* (imperfect *-dir-*) “to put; to make, do” and the Dutch infinitive. All verbal inflectional categories are expressed on *dar*. This kind of strategy for verb insertion is common to CS in many language pairs. Backus (1996b: 211-83) presents an overview and discusses this construction at length concerning Turkish/Dutch and CS generally. The development of MA *dar* “to make; do” into what Backus calls a ‘carrier auxiliary’ is largely parallel to what he describes in the case of Turkish *yap-* “to make; do”. I refer to his work, in particular for an insightful discussion about why an ML verb meaning “to make; do” and the Dutch infinitive are used in verb insertion, and the semantic bleaching that accompanies the use of MA *dar* / Turkish *yap-* with a growing class of embedded verb types.

This method of embedding foreign verbs is well attested for many Turkic, Indo-Iranian and Dravidian languages, among others, and it has also been reported with eastern dialects of Arabic¹. Yet the emergence of an inflection carrying auxiliary in MA is somewhat adventitious, since in western (Maghrebine) Arabic dialects, French and Spanish verbs are inserted by means of attaching the Arabic inflectional pre- and suffixes to the French or Spanish infinitive. The same holds for Italian and English verbs in modern Maltese, which is also a western variety of Arabic.² The *dar* plus

¹ E.g. Anatolian Arabic in contact with Turkish and Kurdish (Vocke & Waldner, 1982: XLIV, 215; Procházka, 1995: 194); Lebanese Arabic/French (Abou, 1962: 65); cf. also Smart (1990: 102-3) on Saudi foreigner talk.

² *Pace* Backus who states that “in Morocco, French infinitives are incorporated in constructions with Arabic non-specialized auxiliaries” (1996b: 215). In the example he quotes from Bentahila & Davies (1983), we actually see that Arabic inflection is attached to the French infinitive. In *tajbqa jconfronter* “he keeps confronting” (B & D’s orthography) the second verb begins with the third person prefix; the final vowel /i/ spelled here as *-er* probably reflects the French infinitive, but in MA/French it marks the imperfect stem; the perfect form “he confronted” ends in /a/, *confront a*. So I gloss *ta-ye-bqa y-confront i* (my orthography) as follows: ASP-3-remain 3-confront·IMPF. The auxiliary *yebqa* in this example is coincidental, and plays no role in verb embedding. Most studies of Arabic/French contact in the Maghreb mention this, with a somewhat varying analysis, but see especially Abbassi (1977: 120, 141-2); Heath (1989: 108-12); Caubet (1993, I: 55-6) on MA/French, Marcel Cohen (1912: 431-50) on Algerian Arabic/Spanish, and Herrero-

infinitive construction was signalled for the first time in Nortier's earlier study of MA/Dutch CS. Nortier counts five occurrences, distributed over two out of the thirteen respondents who contributed to her text corpus (Nortier, 1990: 143, 150; p.c.). In the Nijmegen corpus this construction is far more frequent and widespread. The larger part of this chapter is concerned with a detailed description of the instances of verb insertion in this text corpus. Firstly, the main facts about the construction and its occurrence in the Nijmegen corpus will be presented.

The first example shows how Tense, Aspect, Person and Number, as well as negation are marked on the MA 'carrier auxiliary' *dar*:

- (1) b ş-şehh a buzrud ma ka-t-dir-š *studer-en* mezyan, hè?
 with DEF-reality VOC Bouzroud NEG ASP-2-do-NEG study-INF well QTAG
 "But Bouzroud, you're not studying well, are you?" (Samir)

In monolingual MA the verb *dar* and its dialectal variant *šmel* also occur in periphrastic expressions with verbal nouns, as in (2), from Youssi's (1992) MA grammar, and (3), from the Nijmegen text corpus (see also (79) below). In this context *dar* plus a noun is basically a kind of verbalising construction (Backus (1996b) explains this with respect to Turkish *yap-*). El-Idrissi's 1990 dissertation provides a detailed analysis of this construction in MA.³

- (2) t-šeml-i ši šuma?
 2-do-F INDEF swim-VERBALNOUN
 "Would you like to take a swim?" Moroccan Arabic (Youssi, 1992: 99)
- (3) kifaš ka-t-dir daba t-tešhih dyal dyal d-dfater?
 how ASP-2-do now DEF-correction of of DEF-notebook
 "Now how do you do the correction of the notebooks?" MA (Mimoun)

Muñoz Cobo (1996: 146 ff.) on MA/Spanish. For my interpretation, see Boumans (1995a: 296; 1996: 57-8; and especially 1997). On embedded verbs in Maltese, see Aquilina (1965: 201-5); Mazzon (1993: 200); and Mifsud (1992, n.d. [1995]: 154).

³ El-Idrissi concentrates on the '*dar* plus noun' constructions that can be paraphrased by a single verb. Hence, for his purpose, he does not systematically distinguish between verbal nouns and other nouns in construction with *dar*. Only the former can be said to represent the periphrastic construction; the other cases concern *dar N* in the sense of "to place, put, apply N" or "to make N" which can be paraphrased by a verb derived from this noun. Thus, in *daret l-khul f-šini-ha* versus *kehḥlat šini-ha* "she put on black eye make-up" (El-Idrissi, 1990: 23), for instance, *khul* in the first sentence is not a verbal noun, rather, the verb *kehḥel* in the second sentence is derived from the noun *khul* "black eye make-up" (or the adjective *khel* "black").

In the Nijmegen corpus, many of the monolingual instances of *dar* in a periphrastic expression involve a French noun, as in the following examples from Mustafa. The expressions *dar kunṭak* “to make contact” and *dar ḍumand* “to make a request” possibly originate from analogous French expressions (*faire contact*, *faire une demande*), but note that Mustafa himself does not speak any French.

- (4) dar-u kunṭak mʃa beʃḍ-hũm
 do-PL contact with RECIPROCAL-3PL
 “‘They contacted each other.’ MA (Mustafa)
- (5) der-t ḍumand ʃla ḍ-ḍar
 do-1SG request for DEF-house
 “‘I made a request for housing.’ MA (Mustafa)

While in monolingual MA, *dar* plus verbal noun is not very frequent as a periphrastic construction, in some varieties of MA/Dutch codeswitching it has become a very productive way of incorporating Dutch verbs. The use of the construction is widespread among the respondents of the Nijmegen data corpus, and it is not uncommon in the Moroccan migrant community as a whole, according to my observations. Its use is attested for all four of the Hamadi siblings, as well as for Fatima, Jamal, Mustafa and Zineb, however, it was absent in the contributions by Abdellah, Hayat, Maryam, Mimoun and Warda. The construction is particularly frequent in the Hamadi brothers’ speech; Samir, who participated in most of the recordings, produced 135 tokens, and Abdelkrim and Younes, 29 and 9 respectively. Hocine produced another 11 tokens, while the other respondents produced only a few instances each (Nawal Hamadi 2, Fatima 4, Jamal 4, Mustafa 7, Najib 2, Zineb 2). Since Samir produced so many more tokens of this construction than the other respondents, many details can be investigated just for his CS variety only.

Among the respondents who use this construction we observe different degrees of grammaticalisation of *dar* as an auxiliary that incorporates Dutch verbs. In this process of grammaticalisation and semantic bleaching, *dar* loses its original meaning of “to make, do”. The semantic element of volitionality, that is “to engage intentionally in some activity”, disappears, along with the transitive feature of *dar*. Eventually, *dar* adopts the subcategorisation pattern of the embedded verb. Grammaticalisation is nearly total in Samir’s variety, and less so in that of his younger siblings and the other respondents. This is manifest in the types of verb complementation, as will become clear during the course of this chapter.

The data show a neat distinction between pronominal and lexical verb complements. Personal pronoun complements of the embedded verb are nearly always encoded as suffixes on the MA auxiliary verb *dar*, unless they are simply left unmarked. Lexical complements, on the other hand, are almost exclusively in Dutch, whether they are Direct Objects or prepositional complements. Indefinite pronouns

such as “something” are free forms in either language. Both Dutch and MA indefinite pronouns occur as Objects of embedded verbs. In the case of clausal complements of embedded verbs there is no obvious preference for either language. Likewise adverbial modification: embedded verbs can be modified by an adverb in either language, although the order of the adverb relative to the inserted infinitive is different for MA and Dutch adverbs.

The outline of this chapter is as follows. First we will take a closer look at the embedded verbs that do not take any complement. Then, because of the different grammatical constructions they involve, pronominal and lexical complementation of embedded verbs will be examined separately in sections 2 and 3. After that, I will discuss the instances of clausal complementation that occur in Samir’s section of the data. Adverbial modification is examined in section 5 and some minor patterns of verb insertion that occur in the data will be considered in section 6. Sections 7 to 10 can be read as a summary of the data presented in the previous sections, focusing on four topics. Section 7 evaluates the differences in CS behaviour among the respondents. Here, the main question is to what extent the data presented in this section are representative of MA/Dutch generally. Section 8 takes up the discussion of the grammaticalisation of the verb *dar*. The complementation patterns of embedded verbs are summarised in section 9, as this topic is central to the question of whether embedded words import source language grammatical properties into the matrix language structure, or adopt ML grammatical properties instead. And finally, I will assess to what extent the data on verb insertion can be accounted for within the insertion paradigm of the MSA (section 10).

6.1 Embedded verbs without complement

Verbs that take no complement are distinguished from the cases where obligatory complements of embedded transitive verbs are omitted, cf. section 2.3.

dar “to do, make” is a transitive verb, thus the construction *dar* plus infinitive has one extra argument position in addition to those of the embedded verb. The embedded Dutch infinitive occupies the Direct Object position of the transitive verb *dar*. Examples from different respondents are listed below.

- (6) *ila kūn-t dayr-a-ha u ma metʔekkd-a-š b raš-i, ka-n-dir*
 if be-1SG do·PARTICIPLE-F-3F and NEG convinced-F-NEG with self-1SG ASP-1-do
 twijfel-en
 waver-INF
 “If I would do it [i.e. wear a headscarf] without conviction, I would waver.”
 (Fatima)

- (7) ġir t-dir *opnem-en* wella aš žayy t-dir?
 just 2-do record-INF or what come·PARTICIPLE 2-do
 “You’re just going to make recordings, or what did you come to do?”
 (Jamal)
- (8) bezzaf l-mġarba lli bda-w u dar-u *afvall-en*
 many DEF-Moroccan·PL REL start-PL and do-PL drop·out-INF
 “Many Moroccans started and dropped out.” (Hocine)
- (9) għed f l-likol y-dir *knikker-en*, u kan-u l-wlad temma ġad
 stay in DEF-school 3-do play·marbles-INF and be-PL DEF-boy·PL there still

 y-dir-u *knikker-en*
 3-do-PL play·marbles-INF
 “He stayed at school playing marbles, as the children were still playing
 marbles there.” (Abdelkrim)
- (10) ġla šher ġadi ne-mši n-dir *fitness-en*
 in month FUT 1-go 1-do work·out-INF
 “In a month I will go and do my work out.” (Younes)

With the exception of *opnemen* “to record” in (7), the embedded verbs in the above examples are all inherently intransitive; *opnemen* can also be used intransitively, as in the present context.

The verbs *knikkeren* “to play marbles” in (9) and *fitnessen* “to work out” in (10) happen to be de-nominal verbs themselves. The subsequent cycles of verbalisation and nominalisation can be schematized as follows:

Dutch N	→	Dutch denom. V	→	Dutch verbal N (INF)	→	MA V
<i>knikker</i>		<i>knikker-</i>		<i>knikker-en</i>		<i>dar knikkeren</i>
“marble”		“play marbles”		“playing marbles”		“play marbles”

When the initial noun denotes an activity, like the word *fitness* “fitness training”, rather than an object (*knikker* “marble”), the ‘action noun’ can also function as the complement of *dar* “to do”. See for instance *tdir fitness* in (11), Samir’s reaction to his brother’s remark in (10)

- (11) t-dir *fitness?* fin ġadi t-dir *fitness?*
 2-go fitness·training where FUT 2-do fitness·training
 “You’re going to work out? Where are you going to do your work out?”
 (Samir)

6.2 Pronominal complements of embedded verbs

MA and Dutch differ considerably with respect to the personal pronoun system. In particular, Dutch personal pronouns usually correspond to inflectional categories of the MA verb (although MA has free form pronouns in verbless (copular) clauses, and in a number of other uses that have no close parallel in Dutch). With very few exceptions, personal pronouns as complements of an embedded verb are realised as MA suffixes on *dar*, unless they are simply omitted.

As long as *dar* retains its own transitive feature, the Dutch infinitive assumes the syntactic position of Direct Object (DO) of *dar*. If the embedded verb itself subcategorizes for a DO, it can no longer be encoded as such in the *dar* plus infinitive construction, since there is only one DO position. We can observe two types of solution for overcoming this incompatibility. One is to attach DOs of the embedded verb as Indirect Objects (IO) of *dar*, marked by the preposition *l*.⁴ The other is to leave out the pronominal Object altogether, which results in elliptic utterances. In Samir's CS variant, the embedded infinitive does not seem to occupy the DO position, and *dar* inherits the argument structure of the embedded verb, so that a DO of the Dutch transitive verb is marked as a DO of *dar*.

Also in this section, we will examine a few less frequent patterns involving pronouns, viz. pronominal Objects of PP complements of embedded verbs, double marking of Object pronouns in both MA and Dutch, reflexive pronouns, and so-called indefinite and demonstrative pronouns.

6.2.1 MA Indirect Object suffixes

Direct Objects of the embedded verb are encoded as Indirect Objects of *dar* in the periphrastic construction. This is analogous to the periphrastic construction in monolingual MA. In El-Idrissi's words: "le complément (N1) qui est direct dans [N0 V N1] devient indirect dans [N0 dar Dét V-n prép N1]" (1990: 64, where N0 and N1 stand for Subject and Direct Object respectively, and V-n means "verbal noun"). The periphrastic construction in MA is shown in the following constructed examples from El-Idrissi's dissertation.

- (12) a. l-bulis qellb-u š-šeffar
 DEF-police search-PL DEF-thief

⁴ The terms DO and IO are used here to denote syntactic positions, not semantic roles. The preposition *l* distinguishes the IO from the DO; the IO is distinguished from other prepositional (oblique) complements in that it is suffixed to the verb when pronominal. Negation demonstrates this distinction the most clearly, e.g. *ma gūlt-ha-l-ha-š* "I didn't tell it to her" as against *ma mšit-š m ša-ha* "I didn't go with her".

- b. l-bulis dar-u teqliba l š-šeffar
 DEF-police do-PL search to DEF-thief
 “The police body-searched the thief.” MA (El-Idrissi, 1990: 64)

In (12)a *š-šeffar* “the thief” is the DO of *qellbu* “they searched”; in (12)b it is the complement of *teqliba*, the verbal noun derived from *qelleb* “to search”, and is marked by the preposition *l*.⁵ In MA this rule applies to lexical as well as pronominal complements, while in MA/Dutch CS it is restricted to pronominal suffixes, as we will see.

This process can be clearly demonstrated by one of the examples produced by Mustafa, who is a beginning learner of Dutch. In (13) Mustafa says the MA word for “to fill out (a form)” twice in MA, and then restates his point using the Dutch verb *invullen*. This verb takes a DO complement in Dutch, as does *šemmer* in MA. Now we see that the DO of *šemmer* recurs as the Indirect Object of *dar*, marked by the preposition *l*: *šemmer-t-ha* becomes *dert-l-ha invullen*.

- (13) šemmer-t-ha, šemmer-t l-werqa, der-t-l-ha invull-en
 fill-out-1SG-3F fill-out-1SG DEF-paper do-1SG-to-3F fill-out-INF
 “I filled it out, I filled out the form, I filled it out.” (Mustafa)

Similar examples, where the pronominal Object of the embedded transitive verb is encoded as an IO on *dar*, are produced by Jamal, Hocine, Younes, and Samir (see (107) for an example from Hocine).

- (14) t-dir-li-ya beschuldig-en zešma ana kli-t l-yum?
 2-do-to-1SG accuse-INF EPIST 1SG eat-1SG DEF-day
 “Are you accusing me of having eaten today?” [during Ramadan]
 (Jamal)
- (15) ana ma y-dir-u-l-i-š discriminer-en, (b š-šehh kayn-in ši
 1SG NEG 3-do-PL-to-1SG-NEG discriminate-INF with DEF-reality EXIST-PL INDEF
 drari temmak y-gul-u nta t-dir aanpass-en)
 guy-PL there 3-say-PL 2M 2-do adapt-INF
 “As for me, they don’t discriminate against me, but there are some guys out there who tell me: You should adapt [yourself to Dutch culture].” (Younes)
- (16) ana-ya der-t-l-u ontmoet-en f islamitisch-e bijeenkomst f utrecht
 1SG-EMPH do-1SG-to-3M meet-INF in Islamic-AGR meeting in Utrecht
 “I met him at an Islamic meeting in Utrecht.” (Samir)

⁵ See also example (3) above, taken from the Nijmegen corpus, where the complement of the verbal noun *tešhih* “correction” is marked by a genitive construction.

In the above cases, the embedded verbs *beschuldigen*, *discrimineren* and *ontmoeten* take a DO complement in Dutch. The attested pattern is analogous to what El-Idrissi described for MA *dar* plus a noun derived from a transitive verb. Note that the IO cannot generally be explained by the possible influence of the subcategorisation pattern of the MA translation equivalents of the embedded verbs. Like *invullen* in (13), the embedded transitive verbs *beschuldigen* “to accuse” and *discrimineren* “to discriminate”, for instance, translate as transitive verbs in MA (*ttahem* and *meyyez*); in the context of (16), transitive *ontmoeten* “to get to know” may be translated into MA as *t šerref ũa* or *t šaref m ũa*. Therefore, I assume that the complement of the embedded verb is realised in the IO position because the infinitive itself is perceived as occupying the DO position.

In other cases, however, the Indirect Object could possibly be attributed to the influence of a MA translation equivalent. Take the example of *aantrekken* “to attract” in (17). In Dutch, the attracted person or object is the DO of *aantrekken*, but the IO in (17) is in analogy with the MA verb *šeb* “to please” which marks the attracted person as an IO, i.e., *ka-ye- šeb-l-ek* “it pleases you, you like it”, at least in Samir’s variety of MA (an example is found on p. 203 ex. (87); other MA varieties actually have a DO, *ka-y- šeb-ek*, e.g. Fatima’s (134) on p. 214).

- (17) *šta huwa lli ka-y-dir-l-ek aantrekk-en temma?*
 what 3M REL ASP-3-do-to-2SG attract-INF there
 “What is it that attracts you there?” (Samir)

In still other cases, the pronominal IO of *dar* can be said to simply represent the IO of the embedded Dutch verb or its MA translation equivalent. The argument encoded in this position has the semantic role of RECIPIENT or is typically associated with the IO in MA. Examples of this, like (18), (19) and (20), were produced only by Samir.

- (18) *dir-li-ya samenvatting gev-en!*
 do•IMP-to-1SG summary give-INF
 “Give me a summary!” (Samir)
- (19) *waḥed l-weld ũta-ha-li-ya, dar-li-ya len-en*
 INDEF DEF-boy give-3F-to-1SG do-to-1SG lend-INF
 “A guy gave it to me, he lent [it] to me.” (Samir)
- (20) *dir-li-ya uitlegg-en! dir-li-ya uitlegg-en, wat betekent dat?*
 do•IMP-to-1SG explain-INF do•IMP-to-1SG explain-INF what means this
 Explain to me! Explain to me, what does it mean?” (Samir)

6.2.2 MA Direct Object suffixes

In Samir's variety of MA/Dutch CS, the MA verb *dar* is largely grammaticalized to become a semantically empty auxiliary that carries all the inflectional features when a Dutch verb is inserted. As a symptom of this process, *dar* loses its transitive feature, and the construction [*dar* + Dutch infinitive] no longer involves an increase in valency. The embedded verb no longer occupies the DO position of *dar*, so that the DO position of *dar* becomes available for the DO of the embedded verb.

This development can be demonstrated by comparing two instances of the embedded Dutch transitive verb *repareren* 'to repair' produced by Mustafa and Samir in the same conversation. This verb has an IO complement when used by Mustafa (21), and a DO in Samir's variant (22). (Mustafa omits the initial /r/ in the Dutch word *repareren*: [eparerθ].)

- (21) liʔanna hadak ka-y-kun kbir, ma ġadi y-żi-ši n-dir-u-l-u eparer-en
 because DEM ASP-3-be big NEG FUT 3-come-NEG 1-do-PL-to-3M repair-INF
 "Because these [water tanks] are large, we cannot possibly repair them."
 (Mustafa)

- (22) u hadik te-şleḥ-ha, hadik xeşş-ek t-dir-ha reparer-en?
 and DEM 2-repair-3F DEM must-2SG 2-do-3F repair-INF
 "So you repair this, you have to repair this?" (Samir)

Some further typical examples of Samir's codeswitching variety are listed below.

- (23) ma dar-u-hūm-ş n-nas controler-en, škun ġadi y-dir-hūm
 NEG do-PL-3PL-NEG DEF-people supervise-INF who FUT 3-do-3PL
controler-en?
 supervise-INF
 "The people didn't supervise them; who will supervise them?" (Samir)

- (24) ma bġa-w-ş y-dir-u-ni aannem-en
 NEG want-PL-NEG 3-do-PL-1SG admit-INF
 "They didn't want to admit me [at a school]." (Samir)

- (25) haduk interessant-e mens-en, haduk lli ka-ne-bġi n-dir-hūm
 DEM•PL interesting-AGR person-PL DEM•PL REL ASP-1-like 1-do-3PL
ontmoet-en (baş n-hedṛ-u ʕla l-ʔafkar dyal-hūm)
 meet-INF COMP 1-talk-PL about DEF-idea•PL of-3PL
 "Those are interesting people. Those are the ones I like to meet, so that we can talk about their ideas." (Samir)

The embedded verbs in these examples are transitive in Dutch, and the complements marked on *dar* are DOs. This is apparent from the absence of the IO marker *-l-* or *-li-*, as well as from the distinct 1SG Direct Object form *-ni* in (24) (1 SG IO is *-l-i* ~ *-li-ya*). Note in passing that a lexical Subject NP *n-nas* “the people” comes between the inflected verb *dar* and the embedded infinitive in (23).

While examples (23)–(25) are representative of the majority of inserted transitive verbs in Samir’s variant of MA/Dutch CS, sometimes he also marks DOs of embedded infinitives as IOs of *dar*, in the manner described in the preceding section. His oscillation between the two structures even appears in connection with the same inserted verb: compare *ontmoeten* “to meet” in (25), where the complement is a DO, and in (16), where the complement is a IO. In still other cases Samir leaves out the pronominal complement altogether. His brother Abdelkrim displays the reverse pattern: he omits pronominal complements most of the time, but occasionally marks them as DO suffixes (see (91) and (97)). The omission phenomenon is addressed in the next section.

6.2.3 Omission of Object suffixes

Instead of encoding pronominal complements of embedded verbs as either Direct or Indirect Objects on the auxiliary *dar*, some respondents tend to omit pronominal complements altogether.

One of the two embedded verbs produced by Zineb is an example of this (the other concerned an intransitive verb): *uitlachen* “to laugh at” takes an obligatory DO in Dutch. From the context of this example, it is clear that the addressee is the intended DO, however, this is not marked in the clause.

- (26) *ma ka-n-dir-u-š uitlach-en!*
 NEG ASP-1-do-PL-NEG laugh-at-INF
 “We’re not laughing at [you]!” (Zineb)

Note that in Samir’s variety, this sentence is likely to be rendered as *ma ka-ndiru-k-š uitlach-en*, with the 2SG DO suffix *-k*, and other respondents like Mustafa might produce an IO *-l-ek*: *ma ka-ndiru-l-ek-š uitlach-en*.

Of course, the absence of a DO complement in Zineb’s (26) could also imply that she attributes a different argument structure to the verb *uitlachen*: one that does not require a complement. This account has some plausibility since Zineb spoke French and MA better than Dutch at the time, and *uitlachen* “to laugh at” is not a prototypical transitive verb (the MA translation, for that matter, has a PP complement: *dhek fla*). This is not an appealing explanation for the omission of Object pronouns in Abdelkrim and Younes’ speech, however. While only one token was produced by Zineb, missing Object pronouns are definitely characteristic of the CS variety of the younger Hamadi brothers. It is unlikely that Dutch transitive verbs would repeatedly be given the argument structure of an intransitive verb when appearing as embedded

forms. Moreover, Dutch is the main language of daily interaction for Abdelkrim and Younes and we may assume that the subcategorisation patterns of the common Dutch verbs are well entrenched in their mental lexicon.

The omission is demonstrated in the dialogues between Samir and his younger brothers. In subsequent turns, the same Dutch verb is embedded in MA clauses produced by both interlocutors. While Samir marks the obligatory complements as DO suffixes on *dar*, the younger brothers produce elliptic utterances. A dialogue between Samir and Abdelkrim is reproduced in (27) and between Samir and Younes in (28).

- (27) S *waš eh discriminatie ka-t-dir-ha voel-en wella la?*
 Q er discrimination ASP-2-do-3F feel-INF or NEG
 “Discrimination, do you feel that, or not?”
- A *hmm ana la, ma ka-n-dir-š voel-en*
 er.. 1SG NEG NEG ASP-1-do-NEG feel-INF
 “Er .. I don’t, I don’t feel [it].”
- S *ma ka-t-dir-ha-š voel-en?*
 NEG ASP-2-do-3F-NEG feel-INF
 “So you don’t feel it?” (Samir and Abdelkrim)
- (28) S *dar-u-k nta-ya interview-en wella šewwɾ-u-kūm temmak u šafi?*
 do-PL-2SG 2M-EMPH interview-INF or film-PL-2PL there and enough
 “Did they interview you (SG), or did they just film you (PL) there?”
- Y *la, šewwɾ-u-na u šafi, ma dar-u-š interview-en*
 NEG film-PL-1PL and that’s all NEG do-PL-NEG interview-INF
 “No, they just filmed us, they didn’t interview [us/me].”
- S *fiweq ġadi y-dir-u-ha uitzend-en?*
 when FUT 3-do-PL-3F broadcast-INF
 “When will they broadcast it?”
- Y *had s-simana, la, s-simana ž-žayy-a ġadi y-dir-u uitzenden, op*
 DEM DEF-week NEG DEF-week DEF-coming-F FUT 3-do-PL broadcast-INF on
 NOS laat [ɛnoɛs lat]
 NOS Laat
 “‘This week, no, next week they will broadcast [it] in NOS Laat [a current affairs programme]’.”

- S xεṣṣ-ek te-ṣref fiweq ġadi y-dir-u-ha *uitzend-en*
 need-SG 2-know when FUT 3-do-PL-3F broadcast-INF
 “You need to know when they will broadcast it.” (Samir & Younes)

In these two passages we observe that Samir consistently uses DO suffixes where Abdelkrim and Younes consistently omit them. A counter-example was cited as example (15), where Younes does mark an IO. Likewise, Abdelkrim does use a DO suffix in (91) and (97). On the other hand, Samir also occasionally omits an obligatory Object suffix, like the DO complement of *toepassen* “to apply” in (29) below (the prepositional complement *op zijn tekeningen* will be discussed in section 3.3 below).

- (29) (ka-ye-qra *techniek-en* f l-medraṣa, u f d-ḍar huwa ka-y-dir
 ASP-3-learn technique-PL in DEF-school and in DEF-house 3M ASP-2-do
techniek-en toepass-en.) ka-y-dir *toepass-en op zijn tekening-en*
 technique-PL apply-INF ASP-3-do apply-INF to his drawing-PL
 “He learns techniques at school, and [then] at home he applies [the]
 techniques, he applies [them] to his drawings.” (Samir)

Consider also Samir’s examples (19) and (20). The embedded verbs *lenen* “to borrow” and *uitleggen* “to explain” in these examples require a DO complement in Dutch, while an IO is optional. Interestingly, Samir marks the IO but omits the DO suffix (*dar-li-ya lenen* and *dir-li-ya uitleggen* instead of *dar-ha-li-ya lenen* “he lent it to me” and *dir-ha-li-ya uitleggen* “he explained it to me”). Since the contexts that require both a DO and an IO suffix are not so common, we cannot state that the omission of one of the suffixes is a rule or a tendency in Samir’s CS variety. A counter-example, *tdir-ha-l-ek*, is cited below in (51).

6.2.4 MA pronoun suffixes in MA prepositional complements

In addition to IOs and DOs, there are four tokens of MA pronominal complements within a MA prepositional complement (MA pronouns do not occur as Object of a Dutch preposition). All four instances are analogous, and concern the preposition *m ʕa* “with”.

- (30) ka-t-dir mʕa-hūm *voetball-en*?
 ASP-2-do with-3PL play.soccer-INF
 “Do you play soccer with them?” (Samir)
- (31) ka-ye-hḍer mʕa-k, y-qedd y-dir mʕa-k *filosofer-en* u hadak š-ši
 ASP-3-talk with-2SG 3-can 3-do with-2SG philosophize-INF and DEM DEF-thing
 “He does talk with you, he can philosophize with you and so on.”
 (Samir)

- (32) ma ka-t-dir-š mša-hũm omgaa-n?
 NEG ASP-2-do-NEG with-3PL associate-INF
 “You don’t associate with them?” (Samir)
- (33) ġa-ne-bqa n-dir mša-h omgaa-n enzo, weet je wel, uitgaa-n
 FUT-1-remain 1-do with-3M associate-INF etcetera know you AFFIRM go-out-INF
 enzo
 etcetera
 “I would go on seeing him, you know, go out and so on.” (Fatima)

The MA prepositional phrase is an obligatory complement of the verb *omgaan*, cf. Dutch *omgaan met* “to socialise with”, but the Dutch verbs *voetballen* “to play soccer” and *filosoferen* “to philosophize” do not subcategorise for this type of complement and the MA PP is an adjunct in these cases.

The occurrence of these MA PPs is in accordance with the generalisation that pronominal complements of embedded verbs are in MA, while lexical complements are in Dutch. See also the discussion of example (42) in section 2.6 below. The occurrence of a MA prepositional complement is remarkable however, since embedded Dutch verbs almost exclusively select Dutch PP (or lexical DO) complements, as we will see in sections 3.3 and 3.4 below.

6.2.5 Dutch personal pronouns and double marking

As mentioned, personal pronoun complements are in MA, unless they are omitted. I found only two unequivocal counter-examples to this rule, the Dutch inanimate pronoun *’t* [ət] “it” in (34) and *jullie* “you-PL” in (35). In view of numerous analogous instances of inserted transitive verbs in Samir’s speech, we would expect the pronouns to be marked as suffixes of *dar*. Compare *dar jullie ontmoeten* “he met you (PL)” in (35) with *dert-l-u ontmoeten* “I met him” in (16) and *ndir-hũm ontmoeten* “I meet them” in (25).

- (34) baš y-dir-u eh ’t oploss-en
 COMP 3-do-PL er it solve-INF
 “So that they will solve it.” (Samir)
- (35) (ila ža n-gul-u ši waḥed meḡribi hè, u (..), u ža hna-ya) u
 if come 1-say-PL INDEF INDEF Moroccan QTAG and and come here-EMPH and
 dar *jullie ontmoet-en*, *één van jullie twee*, *maakt niet uit*, (u gal eh
 do you-PL meet-INF one of you-PL two ‘matters not matters’ and say er
 (...), *waš ġa-t-dir-i?*
 what FUT-2-do-F

“If let’s say some Moroccan comes, right? And [he has a good position in Morocco], and he comes here and meets you, one of you both, doesn’t matter, and he says .. [and you like each other, and he wants to marry you, but he wants to live in Morocco], what will you do?” (Samir)

In addition, Samir produced a few cases of double marking of the complement pronoun in both MA and Dutch. Consider the following examples.

- (36) ana ma bġi-t-š n-dir-l-ek *jou over..over..* eh eh n-dir n-gul-u
 1SG NEG want-1SG-NEG 1-DO-to-2SG you con.. con.. er, er 1-do 1-say-PL

overtuig-en van mijn gelijk

convince-INF of my right

“I don’t want to conv.. er, er, let’s say to convince you that I’m right.” (Samir)

- (37) ka-t-dir-li-ya *mij aanwijz-en*
 ASP-3F-to-1SG me point.out-INF
 “She pointed at me.” (Samir)

- (38) dar-u-ni-ya *mij terug..drijv-en, naar me eigen wortel-s*
 do-PL-1SG-EMPH me drive.back-INF to my own root-PL
 “They drove me back to my own roots.” (Samir)

In (36) the MA IO *-l-ek* and the Dutch pronoun *jou* refer to the same DO of *overtuigen* “to convince”. Likewise in (37) the MA IO *-li-ya* and the Dutch pronoun *mij* both refer to the DO of the embedded verb *aanwijzen* “to point out”. In (38) we find the MA DO *-ni-ya* and Dutch *mij* as the complement of *terugdrijven* “to drive back”.

6.2.6 Pronouns in Dutch prepositional complements

This section deals with Dutch prepositional complements of embedded Dutch verbs, in which a pronoun is the complement of the preposition. Examples of this occurred only in the contributions of Samir. We distinguish between pronouns that refer to humans and pronouns that refer to non-human entities, since these involve different constructions in Dutch. The use of pronouns in Dutch PPs is of particular interest in view of the overall tendency for personal pronouns to be in MA.

Starting with the human personal pronouns, only one instance occurs in the data. Example (39) is rather straightforward. Note that the complement precedes the infinitive; we will discuss word order in section 3.1 below.

- (39) ka-y-dir *slecht over jou prat-en*
 ASP-3-do badly about you talk-INF
 “He talks badly of you.” (Samir)

In Dutch, non-human pronouns involve a particular construction which requires some clarification. When a personal pronoun that refers to a non-human entity⁶ is the complement of a preposition, it is replaced by the particle *er* and the order of preposition and pronoun is reversed: *ervoor* “for it, for them”, *ernaast* “next to it/them” et cetera. The independent demonstratives *dit*, *deze* “this, these” and *dat*, *die* “that, those” are treated analogously and are replaced by the locative adverbs *hier* “here” and *daar* “there” respectively: *hiervoor* “for this/these”, *daarvoor* “for that/those”. The two elements of these ‘composite pronouns’ may be separated by an intervening constituent.

When this type of PP is embedded as the complement of an embedded verb in Samir’s speech, we observe the following pattern: the preposition subcategorized for by the verb is present, but the pronominal element {*er*, *hier*, *daar*} is missing. Consider the following examples, in which the missing pronominal element is indicated by a line in the example sentence and between square brackets in the glosses. These markers indicate the positions where the pronominal element would occur according to Dutch syntax in the case of the embedded verb occurring as a clause-final infinitive (cf. the discussion of word order in 3.1). Sometimes two positions would be possible in Dutch. See also (70) and (111) for two further examples.

- (40) u n-dir-u *passage-s* *uit hal-en* u n-dir-u bħal hakda
 and 1-do-PL [it] passage-PL [it] out take-INF and 1-do-PL like this-way
 “And we’ll take some passages out of [it], and we’ll do it this way.”
 (Samir)
- (41) ka-y-dir-u *in trapp-en*
 ASP-3-do-PL [it] in step-INF
 “They step in [it].” (expression for “they let themselves be fooled”) (Samir)

The pronominal element is present in just one case, but here it is rendered in MA. In (42) we find an instance of *hier-voor* “for this” in which the Dutch adverb *hier* is replaced by the corresponding MA locative adverb *hna* “here”. In MA, *hna* never assumes the function of pronoun.

- (42) waš ġadi t-dir hna voor kiez-en?
 Q FUT 2-do here for opt-INF
 “Are you going to opt for this?” (Samir)

⁶ The Dutch 3M and 3PL forms are used for count nouns; *het* ~ *'t* is used to refer to mass nouns and units of discourse (as in *ik weet het* “I know it”, where *het* refers to the content of the previous utterance). Animals are sometimes treated grammatically as humans.

This example once again illustrates the strong preference for MA pronouns. Furthermore it shows the governing nature of the adposition, which selects a particular type of pronoun complement and imposes the order Complement-Preposition, which is totally nonexistent in MA.

6.2.7 Reflexive pronouns

The respondents Najib, Jamal, Younes and Samir insert Dutch reflexive verbs in the *dar* plus infinitive construction (1, 1, 2 and 14 tokens respectively). In the few instances produced by Najib and Younes the reflexive pronoun was simply left out. The verbs *inschrijven* “to register” in (43) and *aanpassen* “to adapt” (44) subcategorize for an obligatory DO or reflexive complement in Dutch. From the context it is apparent that it is the reflexive pronoun which is missing here (see also *aanpassen* in (15) above). The pronoun that would be appropriate in Dutch is indicated between square brackets.

- (43) ka-t-dir ____ *inschrijv-en*, t-dir ____ *inschrijv-en*
 ASP-2-do [2·REFL] register-INF 2-do [2·REFL] register-INF
 “You register, you register [yourself, for a course].” (Najib)

- (44) (*ik ken er heel veel*) zeŋma ma y-dir-u-š ____ *aanpass-en*
 I know PARTICLE very many EPIST NEG 3-do-PL-NEG [3·REFL] adapt-INF
 “I know very many who do not seem to adapt [themselves].” (Younes)

When Samir inserts reflexive verbs, the reflexive pronoun does surface most of the time (10 out of 12 tokens). In addition, the intransitive verb *integreren* “to integrate” occurs twice with a reflexive pronoun, even though it is not reflexive in Dutch. Contrary to DO and IO pronouns, reflexive complements, if any, are always in Dutch with a single exception that I will discuss presently.

Dutch has different sets of reflexive pronouns. The most common ones can loosely be characterized as de-emphasized or emphasized. The de-emphasized forms are {1SG *me*, 1PL *ons*, 2SG/PL informal *je*, formal *u* ~ *zich* and 3SG/PL *zich*}; in the emphasised set *zelf* is added to these forms {*mezelf*, *jezelf* etc.}. The former set is used with both inherently reflexive verbs and reflexively used transitive verbs, while the forms with *zelf* are restricted to reflexively used transitive verbs. In addition there is the more formal 1SG form *mij*, and the substandard set that consists of an unstressed possessive pronoun plus *eigen* “own” {*me eigen*, *je eigen* etc.}.

Samir uses a variety of reflexive pronouns with embedded Dutch verbs, as the following examples demonstrate (cf. also (112) below). Compare *je aanpassen* in (45) with *aanpassen* in Younes’ (15) and (44).

- (45) ġadi t-dir *je* *aanpass-en* *daar?*
 FUT 2-do 2·REFL adapt-INF there

“Will you adapt yourself there?” (Samir)

- (46) ana f 1-lewwel kün-t n-dir *mij scham-en*
 1SG in DEF-first be-1SG 1-do 1SG·REFL be·embarrassed-INF
 “In the beginning I was embarrassed.” (Samir)

- (47) u huma ka-y-dir-u eh *zich gedrag-en, thuis,(zoals ze ___ ook bij*
 and 3PL ASP-3-do-PL er 3·REFL behave-INF at·home like they 3·REFL also at

mij gedragen, fhem-t-li-ya?)
 me behave-INF understand-2M-to-1SG

“And they behave at home the same way they behave in my company, you see?” (Samir)

- (48) waš zeřma ġadi te-xtar 1-meğrib bař eh t-dir *jezelf specialiser-en*
 Q EPIST FUT 2-choose DEF-Morocco COMP er 2-do 2SG·REFL specialize-INF

wella ..?

or

“Will you choose Morocco to specialise in, or .. ?” (Samir)

Samir displays a tendency to overuse the emphasized forms with *zelf* with embedded verbs (6 tokens) and even uses these forms with inherently reflexive verbs like *zich specialiseren* “specialise” in (48). Twice the element *zelf* surfaces while the pronominal element is left out: *zelfontwikkelen* instead of *je ontwikkelen* in (cf. Dutch *je ontwikkelt je* “you develop”). Elsewhere, the reflexive pronoun is lacking (2 tokens, e.g. (82) below).

- (49) ka-t-dir *zelf ontwikkel-en*
 ASP-2-do self develop-INF
 “You develop.” (Samir)

In addition to the instances produced by Samir, there is one from Jamal, showing the (substandard) form *me eigen*, in which the possessive pronoun *me* agrees with the Subject of the clause.

- (50) ma ne-qder-ř n-dir *me eigen concentrer-en*, wella ne-qra ři ħaža
 NEG 1-can-NEG 1-do my own concentrate-INF or 1-read INDEF thing
 “I can’t concentrate or read anything.” (Jamal)

The reflexive verbs add another dimension to the phenomenon of verb insertion. Whereas the insertion of simple infinitives or infinitives plus lexical complements can be described as instances of content word and/or constituent insertion, the reflexive pronouns are function words. Moreover, the selection of the appropriate

reflexive pronoun that is co-referent with the Subject expressed on the MA verb *dar* involves a grammatical procedure in Dutch. This phenomenon constitutes a challenge to the Monolingual Structure Approach because, apart from the function morpheme status of reflexives and the EL word order, it involves the insertion of an element that has no clear counterpart in the ML structure.

MA has reflexive pronouns as well, but the use of reflexives is rather different in each language. Dutch inherently reflexive verbs often correspond to MA mediopassive verbs marked by the prefix *t-* ~ *tt-*. For instance, Dutch *zich inschrijven* “to register”, *zich concentreren* “to concentrate” and *zich gedragen*, “to behave” translate as MA *tsežžel*, *trekkez* and *tserref* respectively. Other Dutch inherently reflexive verbs correspond to simple verbs in MA, e.g. Dutch *zich schamen*, MA *ħšem* “to be ashamed, embarrassed”. MA reflexive forms consist of a body part word, usually *raṣ* “head” (but often *ruh* “soul” in Samir’s variety), with a pronominal suffix, e.g. *raṣ-i* head-1SG “myself”, and they occur with reflexively used transitive verbs, like the Dutch forms with *-zelf*. In addition, there is no one-to-one correspondence of MA and Dutch reflexive pronouns; for instance, Dutch 3SG/PL *zichzelf* corresponds to MA {*raṣ-u*, *raṣ-ha*, *raṣ-hũm* ~ *ruṣ-hũm*} “himself, herself, themselves”.

Finally, we note an isolated and curious instance of a MA reflexive complement of an embedded Dutch verb. In (51) the IO suffix *-ek* is co-referential with the Subject of the same verb *dar*. In ‘ordinary’ MA, reflexive forms like this are not possible; a reflexive form *l-raṣ-ek* would be used instead.

- (51) t-dir-ha-l-ek *nog moeilijk-er maken dan het is*
 2-do-3F-to-2SG even difficult-COMPAR make-INF than it is
 “You’re making it even more difficult for yourself than it already is.” (Samir)

6.2.8 Indefinite and demonstrative pronouns

Compared with personal pronouns the independent indefinite and demonstrative pronouns are morphologically more similar in MA and Dutch. They are free forms in both languages and they are not paradigmatically organised. As Objects of embedded verbs they seem to occupy a somewhat intermediary position between personal pronouns and content nouns. While personal pronouns are nearly always in MA and lexical Objects in Dutch, indefinite and demonstrative pronouns from either language occur in this position, although it involves few instances. These MA and Dutch free form pronouns range with lexical DOs in that they are never marked as IOs as we will see in section 3.

a) Dutch indefinite pronouns

Concerning Objects of embedded verbs, I count three instances of *iets* “something” and three of *alles* “everything”.

- (52) ka-t-dir *iets* *opstek-en*, fhem-ti?
 ASP-2-do something learn-INF understand-2SG
 “You learn something, you see?” (Jamal)
- (53) ʕend-ek ši suʔal, t-gul, zeʕma t-dir-l-u *iets* *vragen*, ma
 at-2SG INDEF question 2-say EPIST 2-do-to-3M something ask-INF NEG
 ye-ʕti-k-š *antwoord*⁷
 3-give-2SG-NEG answer
 “[Suppose] you have a question, you say ..., let’s say you ask him something,
 he won’t give you an answer.” (Samir)
- (54) ma ʕemmɾ-u ma eh eh y-dir eh *iets* *over zijn godsdienst lez-en*
 NEgever-3M NEG er er 3-do er something about his religion read-INF
 “He never reads anything about his religion.” (Samir)

In (54) *iets* is modified by a Dutch prepositional phrase. Also note the MA IO *l-u* in (53). The indefinite pronoun *alles* “everything” also occurs three times: in Hocine’s example (60), in the second line of Samir’s (73) and in the next example.

- (55) ne-qedd n-dir *alles* *opschrijv-en*
 1-can 1-do everything write-down-INF
 “I can write everything down.” (Samir)

b) MA indefinite and demonstrative pronouns

There is only one example of a MA indefinite pronoun and one of a demonstrative pronoun. What makes them so special is that, apart from the Object suffixes on *dar*, these two are the only instances containing a MA Direct Object of an embedded Dutch verb.

- (56) baš t-dir hadak *realiser-en*, xeşş-ek t-žib eh n-gul-u eh
 COMP 2-do DEM realise-INF must-2SG 2-get er 1-say-PL er
 “To realise this, you have to get er, let’s say er ..” (Samir)
- (57) ila bği-t t-dir ši ʔaža *bereik-en*, eh, xeşş-ek te-xdem-l-ha,
 if want-2M 2-do INDEF thing reach-INF QTAG should-2S 2-work-for-3F
 xeşş-ek te-xdem
 should-2SG 2-work
 “If you want to attain something, right, you have to work for it, you have to work.” (Younes)

⁷ The combination of MA *ʕta* “to give” and Dutch *antwoord* “answer” is likely to be a calque from Dutch *antwoord geven* “to give an answer, to answer”.

Note that in (56) and (57) the demonstrative pronoun *hadak* “that” and the indefinite pronoun *ši haža* “something” precede the infinitive.

6.3 Lexical complements of embedded verbs

This section discusses the insertion of ‘lexical’ nominal and prepositional complements of embedded Dutch verbs as distinguished from pronominal and clausal complements. There is a very strong tendency for lexical complements to be in Dutch. This particularly concerns the content morpheme part of such complements; MA determiners of Dutch lexical complements do occur.

Another noticeable characteristic of lexical complements concerns word order: the relative order of the Dutch infinitive and its DO or prepositional complement is always in accordance with Dutch syntax, except in the case of the instances uttered by Hocine. The details about word order will be examined in 3.1 below.

In contrast with what happens to personal pronouns, the Direct Object of an embedded infinitive is never marked as an Indirect Object. That is, we do not find the preposition *l* with the Objects of embedded Dutch infinitives. Even in those rare cases where the Object is in MA or when it consists of a Dutch noun preceded by a MA determiner, *l* is absent. Hence the absence of this preposition cannot be due to its relative unproductivity with Dutch words (cf. Chapter 5, section 1.14). This pattern diverges from El-Idrissi’s description of the periphrastic construction with *dar* in (monolingual) MA, where the complement of the verbal noun is marked by *l* (1990: 64; see the quotation and example (12) on p. 228).

Following the discussion of word order I will examine the different lexical complements divided along two axes: DO (and IO) complements as distinguished from prepositional complements and Dutch complements as distinguished from MA ones.

6.3.1 Word order

A remarkable fact about lexical complements is that, with the exception of the instances uttered by Hocine, the relative order of the Dutch infinitive and its complements is always in accordance with Dutch grammar. In Dutch, non-finite verbs are ‘clause -final’, which means that DO complements precede the non-finite verb and clausal complements follow it (see section 4 below), while prepositional complements occur on either side.

Let me illustrate this with a single example from Abdelkrim. (58) represents the pattern of verb insertion that is attested for Jamal, Fatima and each of the Hamadi siblings. Now compare this to its monolingual Dutch paraphrase in (59). (A finite future tense auxiliary *gaan* “to go” is added to the paraphrase in order to trigger the infinitive form of the main verb *gebruiken* “to use” in clause-final position.)

- (58) *ma ka-t-dir-š echt moeilijk-e woord-en gebruik-en*
 NEG ASP-3F-do-NEG really difficult-AGR word-PL use-INF
 “It [a TV programme] doesn’t use really difficult words.” (Abdelkrim)
- (59) *(ze gaan geen) echt moeilijk-e woord-en gebruik-en*
 they will no really difficult-AGR word-PL use-INF
 “They won’t use really difficult words.” Dutch paraphrase of (58)

Apart from left-dislocation constructions, the complements follow their head in MA,⁸ so that the order Complement-Verb must be attributed to Dutch syntax. In the few instances produced by Nawal, Fatima and Jamal and the numerous ones produced by the Hamadi brothers, no single violation of Dutch word order was found with respect to the relative order of the embedded infinitive and its lexical complement. Hocine, however, places DOs after the infinitive and a single less clear-cut instance occurred in Najib’s utterances. As for the remaining respondents, the data corpus does not contain any instance of an embedded verb plus a lexical complement.

Let us take a closer look at Hocine’s codeswitching variety, which diverges so markedly from that of the Hamadi siblings, Fatima and Jamal. Hocine produced four instances of an embedded infinitive plus a lexical DO complement and in each case the order is Verb-Object, that is, the reverse of that of the other respondents:

- (60) *ġa-ne-mši-w l l[?]-bibiotheek, n-dir-u verzamel-en alle alle alle alle eh*
 FUT-1-go-PL to DEF[?]-library 1-do-PL collect-INF all all all all er
alles lli ġend-na
 everything REL at-1PL
 “We’ll go to the library and we collect all, all, all er .. everything we have.”(Hocine)
- (61) *melli ka-t-dir beheers-en dik taal*
 when ASP-2-do master-INF DEM language
 “When you master that language (..)” (Hocine)
- (62) *(A. B. is iemand die echt beheerst __ arabisch-e taal,)*
 [a name] is someone who really masters DEF•C Arabic-AGR language
volgens mij ka-y-dir beheers-en __ arabisch-e taal kteṛ men
 according•to me ASP-3-do master-INF [DEF•C] Arabic-AGR language more than

⁸ Although particularly ‘light’ PP complements consisting of a preposition and a pronominal suffix can be fronted to precede the governing noun or verb, e.g. *lli nebġi-h, ŋli-h nġir* ‘the one I love, I defend him’ (from *Kelma waḥda*, a song by Najat Aatabou). See Lalami (1996).

l-huḷandiya, ja ik bedoel ka-y-dir echt beheers-en de arabisch-e
 DEF-Dutch yes I mean ASP-3-do really master-INF DEF·C Arabic-AGR

taal, hoe dan ook

language ·in·any·case

“A. B. is someone who really masters the Arabic language. In my opinion, he masters Arabic more than Dutch. Yes, I mean, he really masters the Arabic language, in any case.” (Hocine)

Hocine’s divergent Verb-Object order may be related to his imperfect mastery of Dutch syntax. In the recorded material Hocine typically fails to apply the verb-final word order to finite verbs in Dutch subordinate clauses, as exemplified by the first line of (62). Here we see the Dutch relative clause *die echt beheerst arabische taal* in which the Object follows the finite verb, which is incorrect. (Moreover, the obligatory definite article is lacking in the Object NP.) On the other hand, Hocine usually correctly places infinite verbs (infinitives, participles) in the clause-final position in monolingual Dutch clauses.

To Hocine’s instances of Verb-Object order for embedded verbs and their complements we may add the following isolated example produced by Najib, in which the Object is in MA:

- (63) kifaš ka-t-fekker, ka-t-dir interpreter-en, matalen, ş-şala u had
 how ASP-2-think ASP-2-do interpret-INF for·example DEF·prayer and DEM

l-masaʔil d-dini⁹

DEF-matter·PL DEF-religious

“How do you think, how do you interpret, for instance, the prayer and these religious matters?” (Najib)

In (63) the DO of *interpreteren* comes after the infinitive. This is a less clear-cut example of Verb-Object order because of the intervening pauses and the adverb *matalen* “for example”. An alternative analysis of this example would assume an omitted pronominal DO suffix on *ka-t-dir* and consider the clause-final NPs as right-dislocations (“how do you interpret it, for instance ..”). Najib produced no more than two instances of verb insertion so we cannot establish whether he applies the order Verb-Object like Hocine or tends to omit pronominal Object suffixes like Abdelkrim and Younes.

⁹ This adjective lacks an agreement suffix. In this particular case, we would expect the inanimate plural noun to trigger feminine singular agreement: *had l-masaʔil d-dini-ya*.

6.3.2 Dutch lexical Direct Object complements

Lexical DOs of embedded verbs occur with Hocine, Fatima and Jamal and each of the Hamadi siblings. It has already been noted that there is no exact parallel in monolingual MA: MA verbal nouns that enter into the construction with *dar* can have a complement marked by the IO marker *l-* or another preposition, but cannot have a DO.

In many cases the embedded verb and its lexical Object form a more or less idiomatic collocation in Dutch. The category of collocations cannot be sharply delimited although some Verb-Object collocations can be distinguished on syntactic grounds, in particular with regard to the distribution of articles. I will return to this presently.

Dutch lexical Objects may be classified either as embedded nouns or adjective-noun combinations, or as embedded NP constituents. Remember that, although single nouns and adjective noun combinations may constitute full Dutch constituents in certain contexts, they are not considered as sufficient evidence for considering constituent insertion as a type of codeswitching behaviour.

In (64)-(68) the Dutch complement is considered as an inserted noun or an adjective-noun combination.

- (64) n-qedd n-dir *bewijz-en gev-en*
 1-can 1-do evidence-PL give-INF
 “I can provide evidence.” (Samir)
- (65) ma ne-qḍer-š eh n-dir eh *diep-e gesprekk-en voer-en, weet je*
 NEG 1-can-NEG er 1-do er deep-AGR conversation-PL conduct-INF know you

wel
 AFFIRM
 “I can’t er .. carry out deep conversations, you know.” (Jamal)
- (66) n-dir-u *pauze houd-en?*
 1-do-PL break take-INF
 “Shall we take a break?” (Samir)
- (67) t-dir ši, t-dir *toekomst creër-en* temmak
 2-do INDEF 2-do future create-INF there
 “You er, you create a future over there.” (Younes)
- (68) Šebdelkrim y-dir *wedstrijd voetball-en*
 Abdelkrim 3-do game play*soccer-INF
 “Abdelkrim was going to play a soccer game.” (Nawal)

In some of these cases, the complement would count as a full constituent in a corresponding Dutch sentence, because a determiner is not required. Such is the case

for indefinite plural nouns as in (58), (64) and (65) above. In addition, a determiner is not required in certain idiomatic collocations like *pauze houden* “take a break” in (66). Such idiomatic expressions are not readily distinguishable from compound verbs (Geerts et al., 1984: 150, 514).¹⁰ Elsewhere, however, the Dutch equivalent would require a determiner, e.g. the indefinite article in the case of indefinite singulars like in (67) and (68): *een toekomst*, *een wedstrijd*. On the other hand, a MA zero article could be assumed in the above examples, in which case the complements would be considered MA constituents in which Dutch content words are inserted. This is a viable possibility in view of the fact that Dutch Object nouns are occasionally determined by a manifest MA determiner, examples of which will be discussed in the next section.

Example (69) below is different in this respect. The noun *vertaalvariant* “translation variety” refers to one of several curricula programmed for the study of Arabic at the University of Nijmegen. There is just one such “translation variety”, which is uniquely identifiable to the interlocutors, so we can assume that a definite article (in either language) is lacking here.

- (69) ka-n-dir *vertaal-variant kiez-en*
 ASP-1-do translation-variety choose-INF
 “I choose the ‘translation variety’.” (Samir)

In the following examples the DO falls under one of the categories of embedded NPs that were discussed in section 4 of Chapter 5. We might consider the embedded verb together its lexical complement as an embedded constituent; I will return to this in Chapter 10.

Embedded NP constituents that include a Dutch definite or indefinite article are rare in the corpus as a whole, in particular if we consider their relative frequency as compared to other NP types in monolingual Dutch (cf. Ch. 5 section 4.1). As complements of inserted verbs, Dutch NPs that include an article do not appear to occur more frequently. In other words, the context of the Dutch verb does not make Dutch articles more ‘acceptable’ inside MA matrix clauses. A few examples are attested, however: in the third line of Hocine’s example (62) above we find the Object NP *de arabische taal*, which is marked by a Dutch article. It may be recalled that this type of NP insertion appears relatively frequently in Hocine’s contribution to the data. Samir also produced a few cases of Dutch articles in complements of an embedded verb, consider *de vertaling* “the translation” in (70) (the omission of the pronominal complement of the preposition *naast* in this example is discussed in section 2.6) and the NP *de vraag* “the question” in (92) below.

¹⁰ Likewise, *afscheid nemen* “bid farewell” in (84), *strijd leveren* “struggle” (85), *rekening houden* “bear in mind” (86) and *stage lopen* “do a work placement” in (113) are idiomatic expressions in which the Object noun does not receive a determiner in Dutch, although a determiner is sometimes possible in specific contexts.

- (70) der-t *de vertaling* __ *naast legg-en, nederlands-e vertaling*
 do-1SG DEF·C translation [it] next·to lay-INF Dutch-AGR translation
 “I put the translation next to [it], [the] Dutch translation.” (Samir)

In (71) and (72) we find a noun determined by a possessive pronoun (*m’n moeder, mijn hobby’s*). There are four instances of such nominal constituents as Objects of an embedded verb (Samir 3 tokens, Fatima 1).

- (71) *ja, ka-n-dir m’n moeder help-en, natuurlijk, de huishouding doe-n, u*
 yes ASP-1-do my mother help-INF of·course DEF housekeeping do-INF and

ka-n-xeyyeṭ meestal u lezen, lezen ook
 ASP-1-sew usually and read-INF read-INF too
 “Yes, I help my mother, of course, do the housekeeping, and I usually sew, and reading, I read too.” [response to “What do you do in your spare time?”]
 (Fatima)

- (72) *Ṭend-i l-weqt bezzayef baš n-dir eh mijn hobby-s uitvoer-en*
 at-1SG DEF-time much COMP 1-do er my hobby-PL perform-INF
 “I’ve got lots of time to carry out my hobbies.” (Samir)

As for the phrase *de huishouding doen* “to do the housekeeping” in (71), this may either be analysed as a complement of *ka-ndir*, in coordination with *m’n moeder helpen* “to help my mother”, or as an independent elliptical clause, similar to *lezen* “to read” at the end of the utterance.

The next category of Dutch nominal constituents consists of nouns determined by a quantifier. One such NP occurred as the DO of an embedded verb: *veel moeilijke woorden* “many difficult words” in the first line of (73) is determined by the quantifier *veel*.

- (73) *huma ka-y-dir-u eh veel moeilijk-e woord-en gebruik-en, ontzettend*
 3PL ASP-3-do-PL er many difficult-AGR word-PL use-INF terribly

moeilijk te lez-en, i baš y-dir-u ehm alles opschrijv-en
 difficult to read-INF just COMP 3-do-PL er everything write·down-INF
 “They use many difficult words, terribly hard to read, just in order to write down everything.” (Samir)

The last type of embedded NP complements concerns pronouns. It should be recalled that, because free form pronouns have a distribution similar to other NPs, they are classified as constituents rather than content words (Ch. 2, section 2.2.2; Ch. 5, section 4.4). Dutch personal pronouns as Objects of embedded verbs have been discussed in section 2.5 of this chapter. Such Objects turned out to be very rare: five

instances were produced by Samir, three of which involve the reiteration of a MA pronominal suffix by a Dutch free form pronoun. Besides these five complements of embedded verbs, no Dutch personal pronoun is inserted in a MA clause (Ch. 5, section 4.4). Indefinite pronouns more resemble content nouns. They are free forms in both languages and they are not paradigmatically organised. They are more freely inserted in MA clauses than personal pronouns, both as Objects of embedded verbs, and elsewhere (see Ch. 5, section 4.4). As shown in section 2.8 above, the Dutch indefinite pronouns *iets* “something” and *alles* “everything” each occur three times as the DO of an embedded verb.

In conclusion, embedded Dutch verbs seem to ‘trigger’ exclusively Dutch lexical DOs, however these lexical Objects are not very different qualitatively from the embedded nouns, adjective-noun combinations and NP constituents that are found in other positions in MA matrix clauses (cf. Chapter 5). For instance, the insertion of personal pronouns and of Dutch NPs that contain an article is rare, and this also holds for DOs of embedded verbs.

6.3.3 MA lexical Direct Object complements

MA lexical complements of embedded Dutch verbs are very few. This contrasts with MA personal pronoun complements on the one hand, and Dutch lexical complements on the other. The only straightforward examples with an entirely Moroccan Arabic DO complement are reproduced in section 2.8 as (56) and (57). In these examples the demonstrative pronoun *hadak* “that” and the indefinite pronoun *ši ḥaža* “something” precede the infinitive, the same as Dutch lexical complements in the CS varieties of Samir and Younes and the other respondents except for Hocine. A less clear-cut instance of a MA Direct Object occurs in Najib’s contribution to the data corpus, see the discussion of (63) above.

Besides the entirely MA Object NPs in (56) and (57) (and possibly (63)), we note four Dutch nouns marked by a MA determiner: Hocine’s *dik taal* “that language” in (61) and *hadak ervaring* “that experience”, *ši vooruitgang* “some progress” and *ši paar* “a couple” in the examples below.

- (74) daʔimen mnin der-t hadak *ervaring opdoe-n*, (..)
 always when do-1SG DEM experience get-INF
 “Ever since I had this experience, (...)” (Samir)

- (75) u t-dir ši *vooruitgang boek-en*, wella la?
 and 2-do INDEF progress achieve-INF or NEG
 “And do you make some headway, or don’t you?” (Samir)

- (76) nta ġadi t-dir ši paar rak-en wella ..
 2M FUT 2-do INDEF couple hit-INF or
 “You will hit a couple [of them], or ..” (Abdelkrim)

The MA determiner in these examples leads to the classification of the Object NP as a MA constituent in which a Dutch content morpheme is inserted. At first sight, MA nominal constituents as Objects of embedded verbs seem to be rare, however the equivocal status of Dutch complement nouns that are not marked by any determiner should be kept in mind. These constitute the majority of the lexical Objects of embedded verbs, and most of the time the absence of a determiner can be explained within MA grammar. The DO complement is one of the contexts where the indefinite ‘zero’ article is common in MA (Caubet, 1993, II: 266). In other cases, the absence of a determiner in Dutch DO complements of embedded Dutch verbs may be related to the attested unproductivity of the MA definite prefix *l-* (as distinguished from indefinite *wahed l-*). This determiner would be appropriate in MA when the DO is definite, or when it is an indefinite mass noun. On the other hand, the composite indefinite article *wahed l-* is also common with DOs in MA (Caubet, 1993, II: 268), so it is striking that this article does not occur as the determiner of Dutch nouns that are the DO of an embedded Dutch verb. Elsewhere, MA *wahed l-* is a recurrent determiner of inserted Dutch nouns, although the prefix *l-* is usually omitted (cf. Ch. 5 sections 1.4 and 1.5).

Finally, whereas entirely MA lexical NPs are rare as Objects of an embedded verb, this apparent incompatibility can be solved in a left-dislocation construction. The fronted MA constituent then recurs as a pronominal suffix on the ‘auxiliary’ verb *dar*. Such constructions occur in Samir’s contributions (subscript _i marks the co-indexed NPs and suffixes):

- (77) bezzaf šhab-i_i der-t-hūm_i verliez-en
 many friend·PL-1SG do-1SG-3PL loose-INF
 “I lost many friends.” (Samir)
- (78) šhal mušiba_i der-t-ha_i ana-ya der-t-ha_i meemak-en!
 how·many disaster do-1SG-3F 1SG-EMPH do-1SG-3F experience-INF
 “How many disasters have I been through!” (Samir)

6.3.4 Dutch lexical prepositional complements

Prepositional complements of embedded verbs are common in the speech of Samir, and the large majority of such PP complements are Dutch. In addition, we find some Dutch PP complements with Abdelkrim, Younes and Hocine. In monolingual Dutch, PP complements (or adjuncts) may either precede or follow the clause-final verb. In

MA, complement PPs typically follow the governing verb or noun as in the examples (3)-(5); (5) is repeated here for convenience:

- (79) der-t ɖumand ʃla ɖ-ɖar
do-1SG request for DEF-house
“I made a request for housing.” MA (Mustafa, example (5) above)

This indicates that PP complements that precede the embedded infinitive display Dutch word order. The choice of the preposition reflects the subcategorisation pattern of the Dutch verb. Consider the following examples.

- (80) u y-dir door de ben-en schupp-en, door je dijben-en en je
and 3-do through the leg-PL kick-INF through your thigh-PL and your

ander-e dijbeen, y-dir doorschiet-en
other-AGR thigh 3-do shoot-through-INF
“And he kicks through your legs, through your thighs and your other thigh; he shoots through.” (Abdelkrim)

- (81) huma bɣa-w y-dir-u-ni-ya ergens stopp-en, u dar-u-li-ya naar
3PL want-PL3-do-PL-1SG-EMPH somewhere put-INF and do-PL-to-1SG to

___ LTS [ɛlteɛs] stur-en
[DEF.C] LTS send-INF
“They wanted to put me somewhere, so they sent me to the LTS [lower technical formation]”. (Samir)

- (82) (ana gūl-t f ɾaɣ-i zeʃma t-kun ʃend-i zeʃma muhimm žurnal¹¹)
1SG say-1SG in self-1SG EPIST 3F-be at-1SG EPIST anyway newspaper

baš eh n-dir eh ___ op de hoogte stell-en van het nieuws
COMP er 1-do er [me·REFL] on the height place-INF of the news
“I said to myself: at least I will have a newspaper to acquaint [myself] with the news.” (Samir)

In (82) the adverb *ergens* “somewhere” and the PP *naar LTS* “to the LTS” express the goal of a motion (‘allative’) as complements of the verbs *stoppen* “to put” and *sturen* “to send”. In Dutch, place adverbs like *ergens* can have a locative or allative meaning without further marking. In (81) the first PP *op de hoogte* forms an idiomatic expression with the verb *stellen*, the expression as a whole meaning “to acquaint (s.o.,

¹¹ Note the agreement failure: the masculine noun *žurnal* is the Subject of the feminine verbal form *tkun*.

with s.th.)”. It has the PP *van het nieuws* as its complement. The reflexive pronoun that would be required in Dutch is lacking in this example. See (29), (38)-(42), (70) and (111) for further examples.

The instances produced by Samir and Abdelkrim reflect common subcategorisation patterns of embedded Dutch verbs. In these cases verb and preposition form a collocation. In the example produced by Hocine, on the other hand, the ‘instrumental’ argument is not an obligatory complement.

- (83) nta ka-t-dir *prober-en met jouw life*
 2M ASP-2-do try-INF with your life(English)
 “You try [to manage] with your life.” (Hocine)

The verb *proberen* “to try” normally subcategorises for a DO or clausal complement in Dutch. In an alternative analysis, the phrase *met jouw life* could be interpreted as an adjunct (instrumental) PP where the complement of *proberen* is simply lacking.

6.3.5 MA lexical prepositional complements

Besides the instances of *m ʃa* “with” plus pronominal suffix discussed in section 2.5 there are only three instances of MA prepositional complements of an embedded verb. Two of these occur in Samir’s speech and in these cases a clause is the complement of the MA preposition. This is all the more conspicuous if we consider that all the lexical complements of embedded verbs, whether DOs or marked by a preposition, are Dutch. The third instance of a MA prepositional complement, produced by Abdelkrim, is the only exception in the Nijmegen corpus.

- (84) kan-et ʃend-i wahed l-weqt kũn-t eh eh der-t *afscheid nem-en* n-gul-u
 be-3F at-1SG INDEF DEF-time be-1SG er er do-1SG farewell take-INF 1-say-PL

dyal eh *marokkaan zijn*
 of er Moroccan be-INF

“I had a period in which I was er er .. bade farewell let’s say, to er being Moroccan.” (Samir)

- (85) (huwa b ruḥ-u daʔimen ka-y-dabez mʃa d-din, maši ka-y-dabez
 3M with self-3M always ASP-3-fight with DEF-religion NEG ASP-3-fight

zeʃma) ka-y-dir *strijd lever-en* mʃa : waš ka-n-tiq wella ma
 EPIST ASP-3-do fight produce-INF with Q ASP-1-believe or NEG

n-tiq-š, waš kayen llaḥ wella ma kayen-š, waš eh (...)

1-believe-NEG Q EXIST God or NEG EXIST-NEG Q er

“He himself always struggles with the religion, not that he fights, it’s like he struggles with: ‘Do I believe, or not? Does God exist or not?’ (..)” (Samir)

- (86) ka-n-dir *rekening houden* mʃa [unintelligible] u l-qraya
 ASP-1-do account keep-INF with ? and DEF-education
 “I take into account [unintelligible] and my education.” (Abdelkrim)

Although the preposition itself is in Arabic, the subcategorisation patterns of the embedded Dutch verbs are respected. MA *dyal* in (84) stands for Dutch *van*: *afscheid nemen van* “to bid farewell to”. This is again a context where *van* has the connotation of SOURCE, ORIGIN and this meaning is transferred to MA *dyal* because of the overlap of *van* and *dyal* in other usages, notably possessives. A very similar case, *afgeleid dyal* “derived from” in example (111), is discussed in section 3.2 of Chapter 5. As for *mʃa* “with” in (85) and (86), it probably stands for the Dutch preposition *met* “with”: *strijd leveren met* “to struggle with” and *rekening houden met* “to take into account”. The use of *mʃa* in these contexts is not so markedly un-Arabic as is *dyal* in (84), though; compare the (correct) MA form *ka-ndabez mʃa* “I struggle with” in the first line of (85).

6.3.6 Predicative complements

The 136 Dutch infinitives that occur as embedded forms in Samir’s utterances also include two instances of the copular verb *worden* “to become”. In addition, there are two instances of the transitive verb *maken* “to make” which takes the same kind of predicate as *worden*, in addition to a DO. (In this use *maken* can be considered the causative form of the copular verbs *zijn* “to be” and *worden* “to become”.) The predicate complement of these verbs is Dutch in each case. Two instances are reproduced here. The two other ones are cited above in (51) and below in the second line of (96). Note that in (87) *maken* has a lexical DO which is also in Dutch.

- (87) ġadi n-dir-u *probleem nog moeilijk-er* mak-en
 FUT 1-do-PL problem still difficult-COMPAR make-INF
 “We will make the issue even more complicated.” (Samir)
- (88) (ka-t-dir *zelf ontwikkel-en*) b ʃ-ʃeħħ ma t-qedd-š t-dir
 ASP-2-do self develop-INF with DEF-reality NEG 2-can-NEG 2-do

intelligent-er word-en
 intelligent-COMPAR become-INF
 “You develop, but you cannot become more intelligent.” (Samir)

Samir rephrases his statement in (88) in the same passage, this time using a MA copula verb:

- (89) ma t-qedd-š t-welli *intelligent-er*
 NEG 2-can-NEG 2-become intelligent-COMPAR
 “You cannot become more intelligent.” (Samir)

6.3.7 Complements of auxiliary verbs

Two Dutch auxiliary verbs were embedded. These take an infinitive complement in Dutch. Both instances concern the causative auxiliary *laten* “to let”.

- (90) der-t me in lat-en schrijv-en
do-1SG me·REFL register· let-INF ·register-INF
“I had myself registered.” (Samir)

- (91) xeşş-ek t-dir eh vakk-en ma te-şref-t-hüm-ş¹² mliḥ, xeşş-ek t-dir-hüm
must-2SG 2-do er subject-PL NEG 2-know-2M-3PL-NEG well must-2SG 2-do-3PL

lat-en uitvall-en

let-INF fall·out-INF

“You have to er, the subjects you don’t know well, you have to drop them [i.e. from your school curriculum].” (Abdelkrim)

6.4 Clausal complements of embedded verbs

Abdelkrim and Samir are the only respondents who inserted Dutch infinitives that have clausal complements. There are three instances of a Dutch, and three of a MA complement clause. Consequently, there is no indication that clausal complements tend to be in either language, contrary to what we have seen regarding personal pronouns and lexical DO or prepositional complements.

6.4.1 Dutch clausal complements

All three tokens are reproduced below. The complements are embedded questions, introduced by the complementizer *of* “whether”. In (92) the Dutch clause can actually be regarded as the complement of the noun *vraag* “question”, although *vraag* and *stellen* form a set expression. Note that *vraag* is determined by the Dutch definite article, which is appropriate according to Dutch grammar, since the complement clause renders the noun definite. However, Dutch articles are rarely embedded in MA clauses.

- (92) u ka-y-dir de vraag stell-en of zij wel een ziel heeft
and ASP-3-do DEF·C question pose-INF whether she AFFIRM a soul has
“And he poses the question whether she does have a soul.” (Samir)

¹² This verb form contains a prefix *t-* of the imperfect paradigm as well as a suffix *-t* of the perfect paradigm. Generally, such verb forms do not exist in MA, but a similar form occurs in the common expression *ma ne şref* “I don’t know”.

- (93) ana kŭn-t n-dir *twijfel-en* of ik ook *psychologie* zou gaan
 1SG be-1SG 1-do doubt-INF whether I also psychology would go

studerén

study

“I was in doubt whether I would study psychology as well [i.e. like you].”
 (Samir)

- (94) f *derde klas* ka-t-dir *kiez-en* of je naar ____ *HAVO* [havo]
 in third form ASP-2-do choose-INF whether you to [DEF·C] HAVO

gaat of naar ____ *VWO* [feweo]

go or to [DEF·N] VWO

“In the third form you choose whether you go to HAVO or VWO [secondary school types].” (Samir)

6.4.2 MA clausal complements

There are three instances of a MA complement clause of an embedded Dutch infinitive: the embedded question in (95) and the ‘purposive’ clauses in (96) and (97).

- (95) baš y-dir-u *onderzoek-en* ntaš aš t-teʔtir-at dyal l-luğa
 COMP 3-do-PL examine-INF of what DEF-influence-PL of DEF-language

l-huḷandiya šend-ha f l-eh f l-luğa dyal-na

DEF-Dutch at-3F in DEF-erin DEF-language of-1PL

“So that they can examine what kind of influence the Dutch language has on our language.”¹³ (Samir)

- (96) n-nas hna-ya l-muslim-in, xešš-ek ta l-mṛa t-eh n-gul-u eh
 DEF-people here-EMPH DEF-Muslim-PL must-2SG also DEF-woman 2-er 1-say-PL er

t-dir-hŭm *forcer-en* baš y-dir-u *nederlands te worden?*

2-do-3SG force-INF COMP 3-do-PL Dutch to become

“The people here, the Muslims, should you, even the women er, let’s say er, force them to become Dutch?” (Samir)

¹³ Literally : “to examine what is the influence of the Dutch language has on our language”. Samir actually gets mixed up in the formulation of this sentence and indicates the relation between “influence” and “the Dutch language” twice: by means of the possessive marker *dyal* and by means of the verb *šend*- “to have”. *šend-ha* is superfluous and cannot be used after the possessed object (*t-teʔtirat*).

- (97) ta huma xeşş-ek t-dir-hũm kans gev-en bař y-dir-u scor-en
 also 3PL must-2SG 2-do-3PL chance give-INF COMP 3-do-PL score-INF
 “Them also, you should give them a chance to score.” (Abdelkrim)

In (96) and (97) we see a MA complement clause introduced by the MA complementizer *bař*. In both cases the MA complement clause happens to contain an embedded Dutch verb itself: the copula *worden* “to become” plus an adjectival predicate *nederlands* “Dutch” in (96), and *scoren* “to score” in (97).

In (97) we again have a verbal expression, *kans geven* “to give an opportunity”, the complement of which can be viewed as really being the complement of the noun. Dutch nouns that express the concept of possibility are more often followed by MA *bař* plus subjunctive mood, as has already been discussed in Chapter 5, section 1.9, on the complementation of embedded nouns. MA *bař* plus subjunctive is treated as equivalent to the Dutch complementizer *om* plus the particle *te* plus infinitive. Compare the Dutch translations of these clauses: (...) *forceren om Nederlands te worden* and (...) *kans geven om te scoren*. Subordination is doubly marked in (96) by MA *bař* plus subjunctive and by the Dutch particle *te*. Double marking is possible in this particular example because a Dutch infinitive happens to be embedded in the MA complement clause. Note that there is no double marking in Abdelkrim’s example. The presence of the particle *te*, and double marking generally, cannot be accounted for in terms of matrix language and insertion of congruent categories since *te* does not fill any slot in the MA matrix structure.

This section is restricted to the complements of embedded Dutch verbs, but I will present further instances of MA clauses as complements of Dutch verbs in Chapter 9 where Dutch is the matrix language, and the MA clause an insertion type.

6.5 Adverbial modification of embedded Verbs

Embedded Dutch infinitives can be modified by either Dutch or MA adverbs. Examples are found in the speech of Samir, Abdelkrim and Hocine but only Samir used MA adverbs with embedded verbs. In Samir’s CS variety at least there is no such articulate preference for either language as is the case for complements. Dutch and MA adverbs differ noticeably with respect to their order relative to the embedded infinitive: Dutch adverbs precede the infinitive thus respecting the word order that applies in monolingual Dutch. With a single exception MA adverbs, on the other hand, are not found between the auxiliary *dar* and the Dutch infinitive, occurring instead in either clause-initial or clause-final position. In Dutch all adverbs may also occupy the clause-initial position, but the position following the infinitive is possible only for adverbs of time, place and modality on the condition that they are de-emphasized (Geerts et al. 1984:1024). Consequently, most MA adverbs that follow the embedded infinitive verb violate a Dutch word order rule.

The different word order patterns can be clearly demonstrated by means of contrasting pairs of Dutch and MA adverbs that have approximately the same meaning and usage. To start with, compare Dutch *echt* “really” with MA *b ʃ-ʃeħħ* “really”¹⁴.

- (98) mnin dar-u *echt* *sch.. scheld-en*, ma dar-u-š *echt* *scheld-en*
 when do-PL really sc.. scold-INF NEG do-PL-NEG really scold-INF
 “When they were really scolding, they weren’t really scolding.”
 (Abdelkrim)

- (99) ma ka-y-dir-ha-š *men-en* b *ʃ-ʃeħħ* ?
 NEG ASP-3-do-3F-NEG mean-INF with DEF-reality
 “He doesn’t really mean it, does he?” (Samir)

Compare also the position of the Dutch adverbial *zelf* and MA *ruh-ek*¹⁵ in the next examples, which mean “on your own” and “yourself” in their respective contexts, although both can be used with either meaning. The Dutch adverbial is not marked for person (which is appropriate in this context, according to Dutch grammar).

- (100) maši mliħ-a t-dir *zelf* *spel-en*
 NEG good-F 2-do self play-INF
 “It’s not right to play on your own.” [in soccer, a team sport] (Abdelkrim)
- (101) waš t-dir *boterham-pje smer-en* *ruh-ek* wella ..
 Q 2-do sandwich-DIM butter-INF self-2SG or
 “Do you prepare your sandwiches yourself or ..?” (Samir)

There is no instance of a Dutch adverb in clause-initial position, although in both languages adverbs do occur in this position. In other words, embedded adverbs always occur adjacent to the embedded verb. In the next two examples the Dutch time adverb *elke keer* “every time” precedes the Dutch infinitive while MA *ši xetṛat* “sometimes” occupies the clause-initial position.

- (102) ka-n-dir-ha *elk-e* *keer* *uitstell-en*
 ASP-1-do-3F each-AGR time put-off-INF
 “I put it off every time.” (Samir)

¹⁴ Formally, or etymologically, *b ʃ-ʃeħħ* is a PP, and therefore its placement after the Dutch infinitive may not be violating Dutch word order after all, since in Dutch nearly all types of PP constituents can come after the non-finite verb. *b ʃ-ʃeħħ* also occurs as a conjunction meaning “but”.

¹⁵ *ruh-* “soul” for “self” and reflexive forms is typical of the Hamadi siblings in the Nijmegen corpus; the corresponding Atlantic Coast Koine form is *raš-* “head”.

- (103) ši xetɾ-at ka-n-dir *vertal-en*, van *nederlands naar arabisch*
 INDEF time-PL ASP-1-do translate-INF from Dutch into Arabic
 “Sometimes I translate from Dutch into Arabic.” (Samir)

Finally, consider the next two examples from Samir, where the same Dutch verb is inserted, the first time modified by a Dutch, the second time by a MA manner adverb.

- (104) waxxa t-dir *heel snel studer-en*, ma ġadi-š t-dir-ha mezyan
 although 2-do very quickly study-INF NEG FUT-NEG 2-do-3F well
 “You may be studying very quickly, but you won’t be doing it properly.”
 (Samir)

- (105) b š-ṣeḥḥ a buzrud ma ka-t-dir-š *studer-en* mezyan, hè?
 with DEF-reality VOC Bouzroud NEG ASP-2-do-NEG study-INF well QTAG
 “But Bouzroud, you’re not studying well, are you?” (Samir)

Some further examples: a) Dutch adverbs (see also *slecht* (..) *spreken* in (39) above, and *meteen opvallen* in the fifth line of (124) hereafter):

- (106) kün-t n-dir *genuanceed* *denken*
 be-1SG 1-do subtly think-INF
 “I used to keep an open mind.” (Samir)

- (107) wella n-dir-u, n-dir-u-l-u zo *pakken*¹⁶: (dya-ek b l-ṣeḥbiya
 or 1-do-PL 1-do-SG-to-3M this-way deal-with-INF of-2SG with DEF-Arabic

n-dir-u ana, u dya-i b l-huḷandīya t-dir-u nta)
 1-do-3M 1SG and of-1SG with DEF-Dutch 2-do-3M 2M

“Or let’s, let’s deal with it this way: (I’ll do yours [part of a presentation] in Arabic, and you will do mine in Dutch.)” (Hocine)

- (108) maši seḥb-l-ek *voetball-en* t-dir _____ *hel-e tijd*
 NEG seem-to-2SG play-soccer-INF 2-do [DEF/C/INDEF] entire-AGR time

pingel-en u beṣda t-dir *scor-en*
 keep-to-the-ball-INF and then 2-do score-INF

“You shouldn’t think: to play soccer (is) you keep to the ball all the time [or: for a long time], and then you score.” (Abdelkrim)

In the latter example a Dutch noun is used adverbially. In Dutch this expression requires a definite or indefinite article (*de hele tijd* “all the time”, or *een hele tijd* “a

¹⁶ The appropriate Dutch verb here would be *aanpakken* “to deal with” rather than *pakken* “to grab, seize”.

long time”) but no article surfaces here. In this respect the adverbial NP is not different from inserted NPs in other grammatical functions.

b) MA adverbs

- (109) *ma dar-ha-š vertal-en gaʃ*
 NEG do-3F-NEG translate-INF at.all
 “He didn’t translate it at all.” (Samir)

- (110) *šuf kifaš n-dir ana werk-en*
 look·IMP how 1-do 1SG work-INF
 “Look how I work.” (Samir)

There are a few counter-examples in the data with respect to the typical word order of MA and Dutch adverbs. The adverbial uses of MA *šwiya* “a bit” are similar to those of Dutch *’n beetje*. In (111) *šwiya* also has the same syntactic position as its Dutch counterpart relative to the inserted infinitive. Compare *šwiya* in (111) and *’n beetje* in (112). (The omission of *er* in (111) is discussed in 2.6.)

- (111) *ka-n-dir ___ šwiya eh ___ over lull-en, over de samenvatting u*
 ASP-1-do [it] a.bit er [it] about talk-INF about DEF·C summary and
küll-ši (..)
 every-thing
 “I talk a bit about [it], about the summary and everything.” (Samir)

- (112) *zeʃma duk l-mğarba ka-y-dir-u zichzelf ’n beetje controler-en?*
 EPIST DEM DEF-Moroccan·PL ASP-3-do-PL 3·REFL a bit control-INF
 “You mean, those Moroccans control themselves a bit?” (Samir)

This is not to say that the position of *šwiya* in (111) is anomalous from the viewpoint of MA syntax. To my knowledge there is no detailed description of the word order aspects of MA adverbs. In the monolingual MA periphrastic construction cited in (3) at the beginning of our discussion of verb insertion we see that the adverb *daba* “now”, which has a discourse related function in this context, comes between *dar* and the verbal noun.

The Dutch place adverb *daar* “there” in (113) occurs after the non-finite verb. Likewise, see *aanpassen daar* in (45) above. *Daar* is in fact one of the adverbs that may occur in this position in Dutch. The MA *temma* “there” is also found in this position.

- (113) nta ka-t-dir *stage* *lop-en daar*
 2M ASP-2-do apprenticeship walk-INF there
 “You’re doing an apprenticeship over there.” (Samir)
- (114) šta huwa lli ka-y-dir-l-ek *aantrekk-en* temma?
 what 3M REL ASP-3-do-to-2SG attract-INF there
 “What is it that attracts you there?” (Samir), same as (17) above

6.6 Minor patterns of verb insertion

The *dar* plus infinitive construction is by far the most common way to insert Dutch verbs in MA clauses. However, Dutch infinitives and verb stems are occasionally inserted without the MA verb *dar* carrying the inflectional affixes. These insertions are not very frequent in the corpus and tend to be accompanied by pauses and hesitations. I counted 14 occurrences of this (Mimoun 5, Hocine 3, Samir 3 and one instance each for Hayat, Jamal and Warda). Note that this phenomenon is relatively frequent with Mimoun who, like Warda and Hayat, does not use the construction with *dar*. This suggests that this respondent experiences a certain need to use Dutch verbs in the context of a MA clause but lacks a productive insertion strategy.

In four cases the inserted infinitives function as nominal constituents in the MA clause. In the following example from Mimoun, for instance, the coordinated infinitives *aantonen* and *tonen* are the complement of the existential marker *kayen*:

- (115) ma kayen-š eh *amour* wella ši ḥaža, hnaya, kayen matalen hmm
 NEG EXIST-NEG er love(French) or INDEF thing here EXIST for instance er..

belangstelling aanton-en ... wella ton-en wella ..
 interest demonstrate-INF or show-INF or
 “There is no love or something, here. There is for instance er .. demonstrate
 interest .. or show interest, or ..” (Mimoun)

Yet most of these embedded infinitive forms occur in the position of MA verbs in the MA clause. All happen to occur in contexts where a prefixed (imperfect) MA verb would be appropriate and in about half of the cases MA inflectional prefixes precede the inserted verb. I first present the examples without MA prefixes:

- (116) u ġadi-n *werk-en volgens schema, precies volgens schema*
 and FUT-PL work-INF according to plan precisely according to plan
 “And we will work according to plan, precisely according to plan.”
 (Hocine)

- (117) n-gul-u matalen l-wžeh ḥzin židden, waxxa matalen huma
 1-say-PL for·instance DEF-face sad very although for·instance 3PL
 y-ḥawl-u, matalen *acter-en, als je begrijpt*
 3-try-PL for·instance act-INF if you understand
 “Let’s say for instance their face is very sad, even though they try for instance,
 to act [i.e. as if everything is alright], if you understand.” (Mimoun)
- (118) walakin fi-h *devoirs*, fi-h matalen wahed *werkstuk* xexš-ni
 but in-3M homework(Fr.) in-3M for·instance INDEF paper must-1SG
inlever-en
 hand·in-INF
 “But there’s homework in it, for instance there’s a paper in it that I have to
 hand in.” (Mimoun)
- (119) ka-y-gul xexš-kūm *begrijp*
 ASP-3-say must-2PL understand
 “He says you have to understand.” (Warda)

The latter example, produced by Warda, is like Mimoun’s (118) with the Dutch verb as complement of the auxiliary *xexš*- “be necessary”, except that she uses the verb stem instead of the infinitive. The distinction between verb stem and infinitive is not so manifest, though, as they differ only in the final schwa, that is, *begrijp* [bəχrɛip] versus *begrijpen* [bəχrɛipə]. Warda’s mastery of Dutch is not perfect and Moroccan learners of Dutch tend to omit word-final schwas generally. Therefore it is possible that [bəχrɛip] represents the infinitive form after all.

In four cases MA inflectional prefixes precede the inserted infinitive. This does not necessarily mean that these prefixes really attach to the Dutch infinitives, since, in monolingual MA, inflectional prefixes are sometimes uttered separately in hesitations when the speaker is looking for the appropriate verb. Hesitation is obvious because of pauses in the flow of speech in (120) and (121) but it is less evident in (122).

- (120) (la, maši *alcoholic*, bḥal ka-y-dexxel bḥal matalen ʕend-ek ši
 NEG NEG alcoholic like ASP-3-enter like for·instance at-2SG INDEF
aspirateur), bḥal *stofzuiger*, ka-t- eh.. *zuig-en*, (*weet je wel*,
 vacuum·cleaner(Fr.) like vacuum·cleaner ASP-3F- er suck-INF know you AFFIRM
iets dat zuigt)
 something that sucks
 [explanation of the meaning of the MA word *sekkah* for heavy drinker]

“(No, it’s not an alcoholic, (it’s) like it takes in, like for instance you have a vacuum cleaner,) like a vacuum cleaner, it er sucks, (you know, something that sucks.)” (Mimoun)

- (121) ka-n-... *op m’n gemak-je krant-je lez-en*, temmak f *brabantlaan*
 ASP-1- on my ease-DIM newspaper-DIM read-INF there at Brabantlaan
 “I read a newspaper at my leisure, over there at Brabantlaan [a street].” (Jamal)

- (122) ma kayen temmak ma t- *opbouw-en*
 NEG EXIST there REL·PRONOUN 2- build·up-INF
 “There’s nothing to build up over there.” (Hayat)

The last example, from Hocine, also involves the verb stem rather than the infinitive form and is preceded by MA verbal prefixes. In this example the verb stem is more perceptibly distinct from the infinitive than in Warda’s (119). In the case of *bewijzen* ‘to prove’, the verb stem distinguishes itself from the infinitive not only in the final schwa vowel but also in the devoicing of the final consonant, i.e. *bewijs* [bəwɛis] versus *bewijzen* [bəwɛizə].

- (123) ʃref-ti ka-t-*bewijs* belli ’t *goed is, snap je?*
 know-2SG ASP-2-prove COMP it right is understand you
 “You see, you prove that it’s right, you understand?” (Hocine)

6.7 Consistency and variation among the respondents

This section evaluates the differences among the respondents with respect to verb insertion. The main question is to what extent the data presented in this section are representative of MA/Dutch generally. The periphrastic construction with *dar* to incorporate Dutch verbs is fairly widespread among the respondents of the Nijmegen corpus so that it can be considered a recurrent feature of MA/Dutch CS. With some respondents the *dar* plus infinitive construction was found to be very productive. On the other hand, the construction is absent amongst certain others whose speech is otherwise characterised by codeswitching (notably Abdellah and Mimoun). The amount of data gathered for each respondent varies considerably. This seriously complicates both the quantitative and qualitative comparison of the different respondents’ speech behaviour, however, a number of generalisations are possible. On a more detailed level of analysis we observe substantial idiosyncratic variation on the one hand, and the recurrence of the same patterns with speakers of very diverse linguistic backgrounds on the other. Not surprisingly the Hamadi siblings, and particularly the Hamadi brothers, are rather consistent among themselves in their linguistic behaviour. In the case of their sister Nawal, we have too few instances of the *do*-construction at our disposal.

To start with the variation among the respondents, the *dar* plus infinitive construction is most frequent in the speech of the Hamadi brothers. Their utterances display the widest range of embedded verb types (intransitive, transitive, reflexive, copulas and auxiliary verbs) and verbal complementation patterns (DO, prepositional and predicative; pronominal, lexical and clausal), and their embedded verbs are modified by both MA and Dutch adverbs. Also, the grammaticalization of *dar* as a device that incorporates Dutch verbs is the most evident in their speech. Abdelkrim and Samir are the only respondents who mark pronominal DOs of the embedded verb as DO suffixes of *dar*, although they do not do so consistently. Abdelkrim and Younes tend to omit Object suffixes more often than Samir while the marking of pronominal DOs as IOs with *l* is attested for Samir and Younes but not Abdelkrim. The construction with *dar* is also quite frequent with Hocine. He is the only speaker who places lexical DOs after the Dutch infinitive, which may be related to his imperfect internalisation of Dutch syntax.

Concerning the similarity of different respondents' behaviour we note the following. With only marginal exceptions personal pronoun Objects are encoded as suffixes of *dar*, along with the MA verbal categories of Tense, Aspect, Gender and Number and Negation. The tendency to encode the pronominal DO of the embedded verb as an IO suffix on *dar* is attested for a very diverse group of respondents: Younes and Samir, who were raised in the Netherlands, Jamal, who arrived at the age of seven, and the more recent arrivals, Hocine and Mustafa. Moreover, at the time of the data gathering, Mustafa was hardly able to keep up a conversation in Dutch while the Berberophone Hocine had less than native fluency in both MA and Dutch.

In contrast with personal pronouns, content word complements are always in Dutch. The Hamadi siblings as well as Jamal and Fatima respect Dutch syntax rules regarding the relative order of the embedded verb and Dutch complements and adverbs.

6.8 Grammaticalization

It is in the *dar* plus infinitive construction that conventionalisation and grammaticalisation in MA/Dutch CS is most apparent. Grammaticalisation, in Kurylowicz's classical definition, "consists in the increase of the range of a morpheme advancing from a lexical to a grammatical or from a less grammatical to a more grammatical status" (1965: 52). I use the term to refer to the process whereby the MA word *dar* "to do" becomes a device that serves to embed Dutch verbs in MA/Dutch CS. Backus (1996b) describes this process in Turkish/Dutch: the Turkish verb *yap-* develops from a main verb "to make, create", taking as its arguments a human agent and a physical object resulting from the agent's intentional action, into an auxiliary "to do, carry out" 'verbalizing' verbal nouns, i.e. Dutch infinitives. Turkish *yap-* loses aspects of its original lexical meaning such as volitionality and transitivity, a process referred to as semantic bleaching. The development of MA *dar* in MA/Dutch largely parallels

the grammaticalisation of *yap-* in Turkish/Dutch. This is evidenced by the types of verbs that occur as the complement of *dar* and the way in which complements of embedded verbs are encoded.

In section 2 of this chapter we discussed the encoding of pronominal complements of the embedded verb. It was argued that the construction in which the Dutch infinitive occupies the syntactic position of DO of *dar* is closest to the periphrastic construction with *dar* and a (verbal) noun in monolingual MA. In this construction any complement of the embedded infinitive is necessarily realised as an Indirect Object or an oblique complement. This is the pattern found with the respondents for whom MA is still the main language of daily interaction, and whose competence in Dutch ranges from very low (Mustafa) to imperfect (Hocine).

The Hamadi brothers, who were raised in the Netherlands and who are certainly more fluent in Dutch than in MA, display the same pattern but, alongside this, we find another pattern in the speech of Abdelkrim and Samir. In this second pattern, which is dominant in Samir's CS variety, the Dutch infinitive no longer occupies the DO position and the MA auxiliary directly reflects the subcategorisation pattern of the embedded verb. That is, the DO of an embedded Dutch verb surfaces as a DO on *dar*. This indicates that for Samir and Abdelkrim *dar* has lost its original transitive character in this context. It has assumed the primarily grammatical function of a carrier auxiliary that is needed for the insertion of Dutch verbs.

The loss of transitivity is accompanied by the loss of volitionality. Both are demonstrated by the fact that Abdelkrim and Samir also use *dar* to embed non-volitional stative verbs like *voelen* "to feel" and *zich schamen* "to be embarrassed" and even copular verbs: *worden* "to become". Semantic bleaching of *dar* goes hand in hand with the productivity of this construction for embedding Dutch verbs and the frequency of its occurrence in the data. Even without a detailed quantitative analysis, it is apparent that it occurs far more frequently with the Hamadi brothers than in the other respondents' contributions. This is not surprising if we consider that the Hamadi brothers have a limited MA vocabulary and thus need to insert many content words from Dutch if they choose to speak MA. (Apart from CS, their lack of MA vocabulary shows itself in hesitations and their frequently asking the interlocutor for vocabulary items.)

Thus, the qualitative differences in complementation patterns reveal different stages of grammaticalisation. I suspect that the respondents also differ in the types of verbs that can be inserted in the *dar* plus infinitive construction, ranging from verbs that assume the volitional control of an actor to verbs that denote a state undergone by an experiencer and, finally, copulas and other auxiliary verbs (see Backus, 1996b: 241-2). However, since the respondents of the first generation produced only a handful of embedded verbs, the data are not sufficient to really confirm this claim. The non-volitional *twijfelen* "to waver" in Fatima's (6) would be a counter-example, on the other hand.

6.9 Complementation patterns

With respect to the complements of embedded verbs we can make generalisations regarding a) the type of complement, that is, the sub-categorization patterns of the verb, b) word order, and c) the language in which the complements are realised.

To start with the latter, we have seen that, with very few exceptions, personal pronoun complements are always in MA and lexical complements are always in Dutch. That the personal pronouns are in MA was to be expected in the framework of the Monolingual Structure Approach, confirming earlier observations on CS within the insertion approaches. Since they are function morphemes, they are not commonly embedded categories, thus they are realised in the ML (cf. Ch. 1, section 3.3 and Ch. 2, section 2.1). The non-occurrence of Dutch personal pronouns in MA clauses is a general feature rather than specific to the complements of embedded verbs (cf. Chapter 5, section 4.4). Apart from this, there is the consideration that Object pronouns are inflectional categories (suffixes) of the MA verb. Whether this consideration plays any role can be examined by comparing the *dar* plus infinitive construction in MA/Dutch CS with similar constructions in other CS varieties. Turkish/Dutch is a suitable candidate for this, as it has a very similar way of embedding Dutch verbs, while Turkish has free form Object pronouns. Backus did not find any Dutch pronominals (in the widest sense of the word) as complements of embedded Dutch verbs in his Turkish/Dutch data (personal communication, 18-09-1997; cf. Backus, 1996b), so the suffix status of the MA Object pronouns does not seem to be a decisive factor here.

The observation that the lexical complements are always in the same language as the embedded verb points to a special relation that holds between the two (cf. Backus, 1996b: 278-80 who shows this to be a recurrent feature in CS generally). This special relation can be stated in terms like ‘collocation’ or ‘fixed expression’. The elements of a collocation co-occur in an utterance either because together they express a single concept and are stored as one lexical entry in the lexicon. The fact that no such special relation was found to hold between the verb and its pronominal or clausal complement supports an explanation in terms of collocations. On the other hand, not all attested verb-object embeddings are recognisable collocations. And, importantly, the ‘collocation hypothesis’ cannot account for the quasi total absence of MA nominal or prepositional complements of embedded verbs. Thus, the observation that embedded verbs take exclusively (embedded) Dutch lexical complements cannot be satisfactorily accounted for at this point and this matter needs further investigation. Note the similarity with the co-occurrence restrictions on EL nouns and ML attributive adjectives and vice versa discussed in the previous chapter.

We also found that the relative order of the embedded verb and its lexical complements respects Dutch word order rules except for the instances produced by Hocine. Firstly, notice that there is no complete overlap of Dutch complements and Dutch word order: in Hocine’s utterances the lexical complements of embedded Dutch verbs are Dutch, but they violate Dutch word order; while the few (partially) MA complements produced by the Hamadi brothers do respect Dutch word order.

For this reason, either phenomenon deserves an independent explanation. To explain this word order phenomenon, in particular the order Complement-Verb that must be attributed to Dutch grammar, we may consider the possibility of regarding the Dutch verb phrase as an embedded constituent. However, the Dutch insertions occurring as complements of embedded verbs are not like the NP and PP constituents found in monolingual Dutch. (Again Backus (1996b: 279) shows this to be a common pattern in CS generally.) In particular, personal pronouns and constituents that contain Dutch articles are seldom inserted. In fact, the Dutch DOs are much like the Dutch nouns, adjective-noun combinations and nominal constituents embedded in MA clauses generally. Insofar as the complement can be identified as a MA constituent, the MSA can describe the pattern as an instance of layered insertion. So *ši vooruitgang boeken* ‘make some headway’ in (75) is assigned the following hierarchical structure: [[*ši* [*vooruitgang*]_{DutchN}]_{MANP} *boeken*]_{DutchVP}. However, the simple omission of Dutch (EL) function morphemes appears to be the dominant pattern, rather than an identifiable MA constituent structure for the lexical complement of the embedded verb. In any case, it turns out that embedded Dutch verbs have a strong influence on the position of their lexical complements, but not on the internal structure of these complements.

Now a major question must be raised concerning the argument structure of the embedded verbs. Do the Dutch verbs bring along their own argument structure or do they receive a matrix language subcategorisation pattern? In the latter case we can view the embedded verb as being assigned to a class of ML verbs with a concomitant subcategorisation pattern, or, alternatively, there is the possibility of the activation of the subcategorisation pattern of a particular corresponding ML verb.¹⁷

The question is not so easy to answer. If we look for the argument structure of a single corresponding verb in MA, we are faced with the problem of identifying this verb with an acceptable degree of certainty. On the other hand, an investigation of MA and Dutch verb classes and the ways in which semantic roles are typically expressed by the argument structure of verbs in either language is clearly beyond the scope of this study.

Having said this, I believe nonetheless that there is convincing evidence in support of the idea that EL subcategorisation patterns are preserved, at least in the speech variety of the Hamadi brothers. This evidence comes primarily from the prepositional complements. In MA as well as in Dutch, various semantic roles may be encoded as DOs, but the selection of prepositions is very language specific, as are the prepositions themselves. When there is a PP complement of an embedded verb, the Hamadi brothers always select the preposition that is expected on the basis of the subcategorisation patterns of the Dutch verb. Even in the few cases where the preposition is a MA lexical item, its selection suggests a Dutch subcategorisation pattern (cf. section 3.4). In this context we also note that Dutch PPs, attested mainly in Samir’s and Abdelkrim’s utterances, occur as complements of embedded verbs,

¹⁷ See the discussion of congruence in Chapter 1, particularly p. 49 ff.

while they seldom occur as complements of MA verbs (see Ch. 7 section 1.1). Therefore, in the *dar* plus infinitive construction, prepositional complements clearly reflect the subcategorisation patterns of the embedded verbs.

In the case of the Hamadi brothers we can expect that the Dutch subcategorisation patterns are well entrenched in their mental lexicon, as they speak this language fluently and indeed better than MA. At the same time it is very possible that respondents who are less fluent in Dutch do assign a MA subcategorisation pattern to an embedded Dutch verb (or even to a Dutch verb in monolingual Dutch clauses). I have not found any clear evidence for this in the data, but I will quote a passage from a dialogue between Samir and Fatima that will illustrate this point.

In the passage quoted hereafter Fatima and Samir discuss the matter of wearing a headscarf in the Netherlands. The headscarf identifies one as a Muslim woman and in Dutch society it attracts the attention of the people in the street and on the bus, for instance. Fatima learns the Dutch verb *opvallen* ‘to attract attention’ from Samir and uses the word herself in a subsequent turn. However, she assigns a different semantic role to the Subject of *opvallen*: in the last line of (124) the Subject of *ydiru opvallen* is not Fatima who wears a headscarf and hence attracts attention, but the people who notice her. Thus, Fatima interprets *opvallen* as meaning ‘to notice’ rather than ‘to attract notice’. She assigns an argument structure to this newly acquired word and, in view of her misinterpretation, this might very well be that of a transitive verb, although the context in (124) does not actually show this (but cf. Dutch *opmerken*, MA *lahed* ‘to notice’, both of which take a DO).

(124) **dialogue between Fatima and Samir**

S u hna-ya ila der-t furana, ġadi t-dir-i *opvall-en* of niet,
and here-EMPH if put-2M[!] headscarf FUT 2-do-F attract·attention-INF or not

layla?

Layla

‘‘And here, if you put on a headscarf, you will attract attention, isn’t that so, Layla?’’ [Fatima answers in Layla’s place]

F wat bedoel je met ‘*opvall-en*’?
what mean you with attract·attention-INF
‘‘What do you mean by ‘opvallen’?’’

S als je .. eh ila ken-ti te-qra-y f ž-žamiša hna-ya, u nti-ya hiya
if you er if be-2F 2-study-F at DEF-university here-EMPH and 2F-EMPH 3F

l-wah... zešma n-gul-u .. *de enige die een hoofddoek draagt*, ġa-t-dir-i
DEF-onl.. EPIST 1-say-PL the only who a headscarf wears FUT-2-do-F

meteen opvall-en gaŋ n-, gaŋ n-nas ġa-y-ži-w
 immediately attract·attention-INF all DEF-pe.. all DEF-people FUT-3-come-PL

y-šuf-u fi-k
 3-look-PL at-2SG

“If you er, if you study at the university here, and you are the only er, let’s say the only one who wears a headscarf, you will immediately attract attention. All the .. all the people will come and look at you.”

F *ja ila kūn-t ġa-n-dir-ha w-ana rašq-a-li-ya, (..) metʔekkd-a ra-na*
 yes if be-1SG FUT-1-do-3F and-1SG please·PART-F-to-1SG convinced-F PRES-1SG

der-t-ha ġa-n-kemmel bi-ha, maakt me niks uit, waxxa y-dir-u
 do-1SG-3F FUT-1-continue with-3F matters·me nothing ·matters even 3-do-PL

opvall-en, (..)
 attract·attention-INF

“Yes.. if I were going to do this, and I wanted it, if I were convinced that I must do it, I would go on with it, I don’t care, even if they attract [pay] attention. (..)”

6.10 Evaluation in the light of the MSA

In this final section of chapter 6, I will evaluate to what extent the findings on verb embedding can be accounted for within the MSA advanced in Chapter 2, that is, as insertions of congruent categories in an ML structure.

We have seen that, in itself, the *dar* plus infinitive construction is a continuation of an existing structure in MA in Morocco. The Dutch infinitive occupies the position of the verbal noun in the MA periphrastic construction *dar* plus (verbal) noun that has been described in El-Idrissi’s 1990 thesis. So Dutch verbal nouns are embedded in a MA matrix structure. We may notice that a periphrastic construction with a verb meaning “to do” exists in Dutch as well. From the viewpoint of the MSA this fact is a mere coincidence and of no relevance for the embedding of Dutch verbs in MA clauses: it is the MA and not the Dutch periphrastic construction that functions as a matrix in which Dutch infinitives are inserted. Moreover, there is no need to invoke the existence of a periphrastic *do*-construction in Dutch in order to explain the attested *dar* plus infinitive in MA/Dutch, since the latter represents a very common strategy in CS. Further note that *dar* plus Dutch infinitive is even found in the utterances of Mustafa, a respondent who hardly spoke any Dutch at the time. It is highly unlikely that the Dutch periphrastic *doen* “to do” has had any influence on Mustafa’s strategy to embed Dutch verbs in MA, all the more so if we consider that periphrastic *doen* has rather specific uses in standard and substandard varieties of Dutch such as the

topicalisation of the main verb and the marking of habitual aspect (Geerts et al., 1984: 564-5; Giesbers, 1984; Cornips, 1994).¹⁸

There is some variation with respect to the complementation types of *dar* plus infinitive in MA/Dutch. In particular, we discussed the more grammaticalised behaviour of *dar* in the CS variety of Samir and his brother Abdelkrim. The grammaticalisation and semantic bleaching of MA *dar* constitute a form of language change which goes beyond the predictive power of the insertion approach, of course, while not at odds with it. After all, the MSA deals with CS between two language systems, however, these need not be the standard varieties of the languages. Real counter-examples are of three types: omission of obligatory elements, the insertion of elements that have no counterpart in the ML structure and violations of the ML word or constituent order.

The omission of function morphemes is a classical problem area for the insertion approach to CS, which emerges here again. Lexical complements of embedded verbs nearly always lack a determiner (sections 3.2 and 3.3 on pp. 246, 249), and sometimes neither MA nor Dutch grammar can account for this. The omission of the pronominal element in Dutch PP complements (section 2.6) is another recurrent feature.

Recapitulating the preceding section on complementation patterns we can observe that embedded Dutch verbs tend to be accompanied by their own, sometimes very specific, EL subcategorisation pattern. The Dutch infinitives select Dutch lexical DO and PP complements and impose EL word order on their lexical complements. The possibility of the order complement-head is not totally excluded in MA, however, the divergent word order behaviour of Hocine (discussed in 3.1) suggests that the order verbal noun-complement is in fact the expected one from the viewpoint of MA (and see also the monolingual sentences (3) and (12)). In view of their EL complementation patterns, embedded Dutch infinitive verbs clearly have no exact match in the MA verbal or nominal word classes. The MSA does not a priori prescribe such an exact correspondence of EL and ML categories, and congruence can only be established post hoc. The essential issue concerns what precisely is embedded. It is apparent that in the case of EL verbs that take lexical complements, this is not just the phonological form of the EL word. (Remember that a possible alternative analysis which views the EL verb together with its complement as an embedded VP constituent was discarded in section 9.) In the summary of chapter 5 it was already concluded that embedded content words may import EL grammatical features into the ML clause. Of course, this observation does not preclude the possibility that, in other cases, the embedded material is indeed confined to the phonological form, while the grammatical properties of the word derive from a congruent ML lemma (as Myers-

¹⁸ Indeed, where Dutch is the ML, Dutch periphrastic *doen* may be used for the embedding of verbs from other languages or varieties. Nuijstens (1962: 154-7) notices this use of *doen* in Dutch dialect/standard language contact. No MA verbs were inserted in Dutch clauses, however.

Scotton proposes, cf. Ch. 1, p. 42) or an ML class of words (see the discussion of example (124) in section 9).

A serious threat to the MSA lies in the insertion of personal pronouns complements (section 2.5), although this is clearly a minor pattern, and especially the reflexive pronouns in the contributions of Samir and Jamal (section 2.7). The insertion of these pronouns runs counter to the overall tendencies in CS, firstly because they are paradigmatically organised function morphemes, and secondly because the reflexive pronouns have no close ML counterpart, and therefore do not fill a slot in the MA clause structure. These phenomena, as well as the instances of double marking (double Object pronouns, and the double marking of the complement clause in (96)) clearly demonstrate that EL grammatical procedures are involved in the assembly of these mixed clauses.

Chapter 7

Dutch Prepositional and Adverbial Constituents in MA

Embedded prepositional constituents (PPs) and adverbs will be treated separately in section 1 and 2 of this chapter, but it will become apparent that embedded PPs and adverbs share a number of functions in the clause, especially in the domain of modality. Section 3 summarizes this chapter.

7.1 Embedded PPs

The embedded PPs found in the data divide into complements and adjuncts. The distribution of complements is governed by the subcategorisation patterns of verbs, nouns or adjectives. Adjuncts on the other hand are optional constituents, their occurrence being subject only to the requirement that the sentence make sense (Andrews, 1985: 89).

Most of the embedded complement PPs are governed by other embedded Dutch words. These have already been examined in the preceding chapters and will not be reiterated here. (See Ch. 5, section 1.9 on complements of nouns and section 3.2 on complements of predicative adjectives and sections 2.6 and 3.4 in Ch. 6 on complements of Dutch verbs.) Excluding the complements of Dutch words, embedded PPs in adjunct position greatly outnumber those in complement position. In this section we will first look at Dutch PPs as complements of MA words. Then the adjunct PPs will be divided into two sections: modal and discourse organizing adjuncts (1.3) and all the other adjunct functions (1.2). Finally, we will examine PPs occurring as predicates of a copula (section 1.4).

7.1.1 Complement PPs

The insertion of complement PPs is interesting because the MA head subcategorizes for a single or occasionally a small set of specific MA prepositions, so that, according to the MSA, a congruence relation must be assumed to exist between the Moroccan Arabic PP and the Dutch one ‘replacing’ it. I found seven tokens of embedded PP complements, all reproduced below: three by Hocine, two by Samir and one each by Jamal and Zineb. Of these at least two seem to be Dutch idiomatic expressions, except that the verb is realised in MA.

Hocine’s example most clearly reflects the subcategorization pattern of the MA verb. In this example the Dutch preposition *over* “about” occupies the position of

MA *ŋa*, compare MA *ŋref* (*ši haža*) *ŋa X* ‘to know (something) about X’ and Dutch (*iets*) *over X weten*.

- (1) *nou A.B. ka-ye-ŋref heel goed over de arabier-en, over de*
 well [a name] ASP-3-know very well about the Arab-PL about the
arabisch-e eh de taal, l-axūr ka-ye-ŋref over de islam
 Arabic-AGR er the language DEF-other ASP-3-know about the Islam
 ‘Well, A.B. knows a lot about the Arabs, about the Arabic er the language;
 the other one knows about Islam.’ (Hocine)

- (2) *dar-u waḥed l-ižtimaḥ hna zeḥma maši over jullie, over eh ki*
 do-PL INDEF DEF-meeting here EPIST NEG about you·PL about er how
[y-]semmi-w-eh over eh de situatie van de nieuwkomer-s uit
 [3-]name-PL-3M about er the situation of the newcomer-PL from
arabisch-e land-en
 Arab-AGR country-PL
 ‘They had a meeting here, not about you, about er, what’s it called, about the
 situation of the newcomers from Arab countries.’ (Samir)

Samir’s (2) is similar to Hocine’s (1) in that the Dutch preposition *over* occurs in the place of MA *ŋa*. Both prepositions introduce the topic of a conversation, a meeting, the topic of knowledge et cetera.¹

The Dutch preposition *naar* in (3) marks the goal of a motion, as the complement of *žaw* ‘they came’. The complement of *naar*, *hulanda*, is in MA again. Because this would otherwise be the only example of a singly embedded preposition, which is moreover a function morpheme, I prefer to analyse this as a case of layered embedding: the MA word *hulanda* is embedded in the Dutch PP *naar hulanda*, while this PP is itself embedded in a MA clause.

- (3) *huma ža-w (als arbeider-s) naar hulanda*
 3PL come-PL as worker-PL to Holland
 ‘They came to Holland as workers.’ (Samir)
- (4) *ne-mši-w (op onz-e gemak-je) daar-heen*
 1-go-PL at our-AGR ease-DIM there-to
 ‘Let’s go there at our leisure.’ (Jamal)

¹ In addition, there is a partial overlap in the notions of spatial location these two prepositions express: MA *ŋa* typically denotes location on, on top of something, and Dutch *over* movement along a path above something.

In (4) *daarheen* “to that” is the complement of the MA motion verb *mša* “to go”. *daarheen* is one of the special forms of PP that occur when an inanimate pronoun is the complement of the preposition (discussed in Ch. 6 section 2.6). Remarkably, in both examples the Dutch complement PP is preceded by an embedded adjunct PP: *als arbeiders* “as workers” in (3) and *op onze gemakje* “at our leisure” in (4).

The remaining two examples betray influence of Dutch subcategorisation patterns. In both cases the selection of the preposition *op* is idiomatic in Dutch. The PPs *op kamers* in (5) and *op vakantie* in (6) constitute a fixed expression with the Dutch verbs *wonen* “to live” and *gaan* “to go” respectively, which are the translation equivalents of the MA verbs *sken* and *mša* used in these examples. So Dutch *op kamers wonen* is “to live in lodgings”² and *op vakantie gaan* is “to go on holiday”. In non-idiomatic uses, we would rather expect the Dutch preposition *in* “in” for location inside a hollow object like a room, while *gaan* “to go” typically selects a preposition that marks a goal (e.g. *naar* “to, toward”) or a source (e.g. *uit* “from”).

- (5) *kamer-s*, *ka-y-sekn-u op kamer-s*, *de meeste ka-y-sekn-u op kamer-s*
 room-PL ASP-3-live-PL on room-PL the most ASP-3-live-PL in room-PL
 “Rooms, they live in rooms, most of them live in rooms.” (Hocine)
- (6) *ne-mši op vakantie*
 1-go on holiday
 “I’ll go [there] on holiday.” (Zineb)

7.1.2 Non-modal adjunct PPs

In the case of non-modal adjuncts, a small number of instances are distributed over relatively many respondents (10 tokens for Samir, 4 for Hayat, 3 for Hocine, 2 for Abdellah and Jamal and 1 each for Younes, Fatima and Najib). In the following presentation of examples I will only loosely systematise the different functions of adjunct PPs. Most examples concern spatial location or metaphorical extensions of this, notably temporal notions.

Six adjuncts literally denote a place. In this context note that the MA preposition *f* is frequently used with Dutch content words to form place or time adjuncts (examples are listed in Chapter 5, p. 217). Note that in the first two examples the

² The plurality of *kamers* is part of the expression, although this is not apparent in the present example since it can also be explained by the plurality of the Subject in the clause. But compare the same expression in Turkish/Dutch, discussed by Backus (1996b: 129):

- (1) *op kamer-s won-en yap-acağ- m*
 on room-PL live-INF do-FUT-1SG
 “I’m going to live on my own.” Turkish/Dutch (Backus, 1996b: 128)

prepositions are modified by adverbs to express the quite specific locations “two hundred kilometres away from” and “diagonally across”.³ Cf. also example (9) in Chapter 9.

- (7) *hetta* *ila* *kūn-ti* *sakn-a* *eh twee honderd kilometer ver van de stad in*
 even if be-2SG live•PART-F er two hundred kilometre far from the city in

marokko of in turkije, *ta-y-gul-u* *ta ntuma* *ʕend-kūm*
 Morocco or in Turkey ASP-3-say-PL even 2PL have-2PL

cultuurverschill-en!

cultural•difference-PL

“Even if you live er two hundred kilometres away from the city in Morocco or in Turkey, they’ll tell you even you have cultural differences!” (Hayat)

- (8) *u f* *tali lqi-t* *waḥed žuž meḡribi-yat*, *gaʕd-in* *lhih* *gaʕ*,
 and in next find-1SG INDEF two Moroccan-F•PL sit•PART-PL over•there all

gaʕd-in *eh tegenover mij, schuin tegenover mij*
 sit•PART-PL er across me diagonally across me

“And then I noticed two Moroccan girls sitting over there, sitting er across from me, diagonally across from me.” (Samir)

- (9) *bijvoorbeeld in nijmegen* *kayn-a* *ɾ-ɾyafa*
 for•instance in Nijmegen EXIST-F DEF-Rifian•PL
 “In Nijmegen for instance there are Rifians.” (Fatima)

- (10) *kūn-t* *xeddam* *bij kopa*, *kūn-t* *eh zeʕma* *snackbar*, *friettent*
 be-1SG work•PART at Kopa be-1SG er EPIST snack•bar chips•stand
 “I used to work at Kopa’s, I was er .. let’s say a snack bar, a chips stand.”
 (Younes)

The following two examples (11) and (12) are metaphorical extensions of spatial location. An example analogous to *op intellectueel vlak* in (12) is also to be found

³ Actually, the ‘locative’ verbs *skən* “to live” and *gʕed* “to sit” in (7) and (8) do require a spatial location to be specified, so the locative PPs in these examples are not really optional adjuncts. But there is no subcategorisation relation as discussed in the preceding section because the choice of the preposition (or adverb) is free, as long as it denotes a spatial location. In these cases the preposition contributes to the meaning of the clause independently from the verb (cf. Andrews, 1985: 91; Geerts et al., 1984: 862). The same reasoning applies to some extent to verbs of motion like *ža* “to come” and *mša* “to go” in (3) and (4) above, where the complement can either indicate the source (*men* “from”) or the goal (*l* “to”).

in Samir's part of the data. Further, note Jamal's manner adjunct *op onze gemakje* 'at our leisure' in (4) above.

- (11) *op de middelbar-e school* $\text{\textcircled{f}}\text{end-h\ddot{u}m}$ *waħed schema*
 at the secondary-AGR school at-3PL INDEF scheme
 'At secondary school they have a scheme.' (Samir)
- (12) *maši la budda baš ne-hdeṛ b l-huḷandīya wella eh.. ġir op*
 NEG absolutely-COMP 1-talk with DEF-Dutch or er just at
intellectueel vlak, zeg maar, xeṣṣ-ha te-fhem
 intellectual level say just must-3F 3F-understand
 'It's not so that I absolutely have to talk Dutch [with her] or er.. Just, let's say, intellectually she has to understand [me].' (Abdellah)

Time and frequency adjuncts were found with Samir (3 tokens) and Najib and Hayat (1 each):

- (13) *ka-ne-mši-w k\ddot{u}ll \text{\textcircled{f}}am-ayn, om de twee jaar*
 ASP-1-go-PL every year-DUAL around the two year
 'We go every two years, every two years.' (Hayat)
- (14) *l-\text{\textcircled{f}}eql dima, matalen be\text{\textcircled{f}}d eh na generatie-s, ġadi ye-bqa ye-kber*
 DEF-mind always for-instance after er after generation-PL FUT 3-remain 3-grow
 'The mind will always, for instance after er.. after generations, it will keep on growing.' (Najib)
- (15) *ila g\text{\textcircled{f}}ed-t hna f huḷanda ġadi daymen voor de rest van je leven,*
 if stay-2M here in Holland FUT always for the rest of your life
ġadi t-kun mferreq m\text{\textcircled{f}}a l-walid dyal-ek
 FUT 2-be separated with DEF-parent's of-2SG
 'If you stay here in Holland, you'll always, for the rest of your life, you'll be separated from your parents.' (Samir)

In the following examples Dutch *om* and *voor* mark a cause (reason) (16), a purpose (17), (18) and a beneficiary (19).

- (16) *waš ka-t-beddel d-din om politiek-e reden-en, kan toch niet?*
 Q ASP-2-change DEF-religion for political-AGR reason-PL can still not
 'Do you change your religion for political reasons? That's impossible, isn't it?' (Jamal)
- (17) *ka-te-qra voor ontspanning (..) ta ana mnin ka-ne-qra l-\text{\textcircled{f}}erbiya,*
 ASP-2-read for relaxation too 1SG when ASP-1-read DEF-Arabic

ka-ne-qra voor ontspanning

ASP-1-read for relaxation

“You read for relaxation. (..) Me too, when I read Arabic, I read for relaxation.” (Hocine)

- (18) la, ka-ne-qra ondertiteling, voor de zekerheid

NEG ASP-1-read subtitles for the certainty

“No, I read the subtitles, just to be sure.” (Abdellah)

- (19) kayn-in mġarba hna-ya lli ka-y-dir-u ši xedma alleen voor

EXIST-PL Moroccan-PL here-EMPH REL ASP-3-do-PL INDEF work just for

zichzelf

themselves

“There are some Moroccans here who do some job just for themselves.” (Samir)

Dutch *met* occurs twice in an embedded PP. This preposition typically denotes accompaniment and instrumentality. While these are not obligatory arguments, not all verbs (or nouns) allow for an instrument role or an accompanying person, so that in these cases the distinction between adjunct and complement is not so evident (cf. Andrews, 1985: 91).

- (20) huwa kan ʕend-u muškila mʕa hadak napolì hè, met zijn club hè?

3M be have-3M problem with DEM Napoli QTAG with his club QTAG

“He had a problem with this Napoli, right, with his club, right?” (Samir)

- (21) nta ka-te-qra l-ktab dya l-huḷandiya met een heel goei-e gevoel

2M ASP-2-read DEF-book-PL of DEF-Dutch with a very good-AGR feel

“You read Dutch books with a very good feel [for it].” (Hocine)

Finally, there are three instances where Dutch *als* “as” marks a capacity. One example has already been quoted: *als arbeiders* “as workers” in (3). A second instance is quoted below (*als bijvak* “as a minor subject” in the third line of (24)) and the third one is reproduced here. Note how the Dutch PP in this example is interrupted by the MA filler word *ngulu* “let’s say”.

- (22) ila ʕta-w-ek zeʕma l-xedma als eh n-gul-u eh scheikundige f

if give-PL-2SG EPIST DEF-work as er 1-say-PL er chemist in

l-meġrib u hna-ya, b-žuž (..)

DEF-Morocco and here-EMPH with-two

“Suppose they offer you work as er let’s say a chemist in Morocco as well as here, both, (..)” (Samir)

7.1.3 Modal and conjunctive adjunct PPs

Several Dutch prepositional phrases serve discourse marking functions. We can distinguish between those that organise sequences in the discourse and those that express the speaker's attitude to what is being said and toward the interlocutors, the so-called modal adjuncts. Except for two examples by Hocine all instances occur in Samir's data.

Adjuncts that express the speaker's attitude, especially those that make his statement sound less determinate, are quite frequent as embedded PPs in Samir's contributions. Examples of this are *naar mijn mening* "to my opinion", *volgens mij* (idem), *in principe* "in principle" and *over het algemeen* "in general". Seven instances were found in Samir's data and Hocine contributed two tokens of *volgens mij* "according to me". (See also the qualifying adverbs in section 2.5 hereafter.)

- (23) *waš te-t-šellem ši ɣaža .. eh volgens jou?*
 Q 2-MP-teach INDEF thing er according-to you
 "Do you learn something, according to you?" (Samir)
- (24) *volgens mij l-luğa ye-mken-l-ek ġa-te-t-šellem-ha bla*
 according-to me DEF-language 3-be-possible-to-2SG FUT-2-MP-teach-3F without

ma te-qra f ž-žamiša, t-dir ši madda ʔūxra f ž-žamiša u
 COMP 2-study in DEF-university 2-do INDEF subject other-F in DEF-university and

t-dir l-šerbiya apart [apar] als bijvak
 2-do DEF-Arabic separately(French) as minor-subject
 "According to me you can learn Arabic without studying at the university; you do some other subject at the university, and you do Arabic separately, as a minor subject." (Hocine)
- (25) *baqi li-ya in theorie nog eh nog twee jaar*
 remain-PART to-1SG in theory still er still two year
 "Theoretically, I still have two years left [to finish my study]." (Samir)
- (26) *la, zešma, zonder vooroordeel, ka-ye-mši-w hadak eh*
 NEG EPIST without prejudice ASP-3-go-PL DEM er
 "No really, without prejudice, those [not completed] go .." (Samir)

There are two discourse structuring PPs that Samir recurrently uses in combination with MA clauses, namely *aan de ene kant* "on the one hand" and its counterpart *aan de andere kant* "on the other hand". They do not necessarily occur in conjunction. Sometimes Samir omits the preposition *aan*. Note also the conjunction *maar* "but" in the third line of (27).

- (27) ana-ya *aan de ene kant* l-walid dyal-i huma, bħal ka-y-ʃerf-u-ni,
 1SG-EMPH at the one side DEF-parent•PL of-1SG 3PL likeASP-3-know-PL-1SG

ana ka-n-werri-hũm ž-žiha lli ka-y-ʃerf-u.. lli ka-y-ʃerf-u-ha
 1SG ASP-1-show-3PL DEF-side REL ASP-3-know-PL REL ASP-3-know-PL-3F

bezzaf, bezzaf, *maar eigenlijk* huma ma ka-y-ʃerf-u-ni-š gaʃ
 much much but really 3PL NEG ASP-3-know-PL-1SG-NEG at'all

ana-ya

1SG-EMPH

“On the one hand, my parents, like they know me, I show them the side which they know, which they know very very well, but actually they don’t know me at all.” (Samir)

- (28) walakin kayen, *aan de andere kant*, kayen waħed l-muškala kbi::r-a
 but EXIST at the other-AGR side EXIST INDEF DEF-problem big•EMPH-F
 “But there is, on the other hand, there is a gigantic problem.” (Samir)

7.1.4 PPs as predicates

MA copula constructions can have three kinds of predicates: NPs, AdjPs and PPs. All of these are also found as embedded constituents (see Ch. 5 sections 3.2 and 4.5 for adjectival and nominal predicates). PPs that function as the predicate of a copular construction must be distinguished from complement PPs. Clearly, a copula does not constrain the selection of a preposition the way other verbs select their complement. On the contrary, the preposition that occurs in a predicate contributes the major part of the meaning of the clause.

Four Dutch PPs occurred as the predicate after a MA copula. Three were produced by Samir, one by Zineb. The predicate *op de hoogte* “informed” in Zineb’s (29) is a fixed expression (it is also found in Samir’s (82) in Chapter 6). In (29) Zineb uses the MA copula *ykun* to express modality (“one should be informed” as against “one is informed”). The latter two examples have ‘zero’ copulas. A zero copula is the default form for the imperfect when no modal aspect is marked.

- (29) l-ʔinsan beʃda y-kun *op de hoogte*, dima ʃaref aš ka-ye-t-gal
 DEF-human at-least 3-be on the height always know•PART what ASP-3-MP-say
 “One should at least be informed, always knowing what’s being said”
 (Zineb)

- (30) ana-ya *uit een islamitisch gezin*
 1SG-EMPH from a Muslim family
 “I’m from a Muslim family.” (Samir)

- (31) had l-Şarab ma ye-htaž-u-š, bħal-na, đ-đimuqratsiya la xaṭer
 DEM DEF-Arab's NEG 3-need-PL-NEG like-1PL DEF-democracy because

l-Şarab eh *ander* eh *van een ander-e ras*

DEF-Arab's er different er from a different race

“These Arabs do not, like us, need democracy because the Arabs are diff.. of a different race.” (Samir)

7.2 Embedded Adverbs and AdvPs

Adverbs constitute a notoriously heterogeneous word class both notionally and grammatically. Adverbs are “modifiers of constituents other than nouns” (Schachter, 1985: 20). This includes adverbs as modifiers of verbs, adjectives, other adverbs and adverbs that modify entire sentences.⁴

Many of the embedded Dutch adverbs modify other embedded categories. We have discussed adverbial modification of Dutch predicate adjectives in chapter 5 (section 3.2) and of embedded Dutch verbs in chapter 6 (section 5). Section 1 of this chapter, on embedded PPs, also contains examples of Dutch adverbs modifying a Dutch preposition: *twee honderd kilometer ver van* “two hundred kilometres away from” in (7), *schuin tegenover* “diagonally across from” in (8) and *alleen voor zichzelf* “just for themselves” in (19).

Embedded Dutch adverbs were not found to modify MA adjectives or adverbs. In this respect, adverbs behave like attributive adjectives: remember that hardly any Dutch adjective is used to modify a MA noun (Ch. 5, 3.1) and inversely, MA attributive adjectives rarely modify embedded Dutch nouns (Ch. 5, section 1.10). In this section we will therefore only be concerned with Dutch adverbs as modifiers of MA verbs, clauses or sentences. Some of these have already been discussed as well: in Chapter 5 we saw Dutch nouns and NPs that function as temporal adverbs in MA clauses (see the list on p. 217).⁵

⁴ To be precise, Dutch time and place adverbs occasionally do modify nouns as in *die troep daar* “that rubbish over there” and *de vergadering gisteren* “yesterday’s meeting”. This use did not occur with embedded Dutch adverbs in our corpus, however, but see examples (45) and (44) in Chapter 10 for MA adverbs serving this function in Dutch clauses.

⁵ Forms like *hiervoor* “for this”, *eruit* “out of it” and so on, which replace prepositional constituents with inanimate pronouns, are also traditionally called adverbs in Dutch grammars (e.g. Geerts et al., 1984: 376). In this description of MA/Dutch CS they are considered PPs and hence discussed elsewhere, in Ch. 6, section 2.6, as complements of embedded verbs, and in section 1 of the present chapter (see the discussion of example (4) in that section).

It is noteworthy how unevenly the embedded adverbs are distributed among the respondents. Nearly all examples come from three speakers: Hocine, Samir and Fatima. Fatima's frequently occurring insertion of Dutch adverbs is surprising since she does not contribute many examples to other categories of embedded material. If we disregard a) the adverbs that modify other embedded categories b) nouns and NPs and numerals as time adverbs and c) the frequent insertion of *bijvoorbeeld* "for example" by Fatima, we find the following distribution of singly embedded adverbs: Fatima 14 tokens, Hocine 12, Samir 6, Maryam 2, Warda 1. If we further consider that Samir participated in nine out of the ten recording sessions while each of the other respondents were involved in just one or two, it becomes evident that Fatima and Hocine are the uncontested champions of single adverb insertion. This is all the more striking in the case of Fatima, since other types of CS, such as embedded nouns for instance, are not very frequent in her speech. Fatima particularly inserts many modal adverbs from Dutch, although she is well represented in other adverb categories as well.

From the viewpoint of the MSA, one question that presents itself is: are embedded Dutch adverbs EL constituents or EL content morphemes? If we disregard adverbially used nouns, all the singly embedded adverbs constitute independent constituents on their own. They are well-formed constituents in Dutch (the EL) and they function as adverbial constituents in the MA (ML) clause. They never combine with MA morphemes to form MA adverbial constituents, unlike embedded nouns, for instance, which can be part of a larger matrix language NP. For this reason, singly embedded Dutch adverbs may be considered EL constituents but they are also EL content morphemes at the same time. Some embedded adverbs are modified by another Dutch adverb, in particular, the degree (or limiting) adverb *heel* "very" as in *heel goed* "very well". Such cases are classed unequivocally as EL constituents. Firstly, their internal structure, viz. the relative order of both adverbs, must be attributed to Dutch; the MA degree adverb *bezzaf* "very" follows the adverb it modifies as in *mezyan bezzaf* "very well". Secondly, degree adverbs form a relatively closed class of morphemes and, as a result, are less liable to occur as embedded morphemes (indeed, they are not found as singly embedded forms) although they can occur within EL constituents. In conclusion, while not all embedded adverbs can be analysed as instances of content word insertion, we can lump together singly embedded adverbs and adverbs modified by a degree adverb as being EL constituents, with the reservation that we disregard adverbially used nouns. In the following discussion I will not distinguish between single adverbs and adverbial constituents.

An intriguing issue with regard to embedded adverbs (or adverbial constituents) concerns their word order properties. While it is generally assumed in the MSA framework that the ML governs the distribution of embedded constituents in the clause, modal adverbs in particular have been shown to display EL word order properties. Examples from various language pairs were discussed in section 2.2 of Chapter 3. It would be worthwhile to investigate this matter for MA/Dutch as well, however, to my knowledge there is as yet no detailed description of word order

properties of the various types of adverbs in MA. Moreover, the word order of adverbs is very complicated. Because of this it is not possible to draw firm conclusions as to whether embedded Dutch adverbs in MA clauses follow Dutch or MA word order. The presentation of the examples is ordered according to the semantic categories as distinguished in Geerts et al.'s reference grammar of Dutch (1984: 376-7).

7.2.1 Place and time adverbs

Place adverbs only occur in combination with embedded verbs in the data: see the verbal complement *ergens* 'somewhere' (82) and the adjuncts *daar* 'there' in (45) and (113) (all by Samir) in Chapter 6 on embedded verbs. Various nouns denoting periods of time are used adverbially, sometimes modified by a number. These have been discussed as instances of noun or NP insertion. See the following examples in Chapter 5: Jamal's *weekend* 'weekend' in (20), Abdelkrim's *minuut* 'minute' in (57), Nawal's *minuut* and *kwartier* 'quarter hour' in (58) and Samir's *de volgende dag* 'the next day' (120).

The other type of time adverbs found in the data has not been mentioned yet: this involves times of the day according to the clock, indicated in numbers. A number of instances of this were produced by Jamal in a discussion about which bus to take. Another example occurs in Hocine's contributions.

- (32) ana gŭl-t-l-ek *drie over* ġadi n-šedd-u ɤ-ɤubis, nta ma bġi-ti-š
 1SG say-1SG-2SG three past FUT 1-take-PL DEF-bus 2M NEG want-2SG-NEG

gŭl-ti-li-ya nta *negen uur*, n-šedd-u daba *drie-en-twintig over*, heeft
 say-2SG-to-1SG 2M nine hour 1-take-PL now three-and-twenty past has

iedereen z'n zin, šafi, ya-k?

everybody his liking enough QTAG-2SG

'I said to you three past we'll take the bus; you didn't want [to go], you said nine o'clock. Now we'll take [the bus at] twenty-three past, so everyone has it his way, OK, right?' (Jamal)

- (33) ka-n-šuf *nieuws*, *altijd om acht uur, om twaalf uur*
 ASP-1-see news always at eight hour at twelve hour
 'I watch the news, always at eight o'clock, twelve o'clock.' (Hocine)

7.2.2 Adverbs of frequency and degree

Frequency adverbs are found with Hocine (4 tokens), Fatima (3 tokens) and Abdellah (1 instance). Five instances are quoted here in (34)-(37), and see *altijd* 'always' in

Hocine's (33) above and Fatima's *soms* "sometimes" further below in (59). (A third instance of *altijd* "always" by Hocine is not reproduced in this presentation.)

- (34) l-walid ta huwa zo nu en dan ye-ḍreb
 DEF-father also 3M so now and then 3-phone
 "My father calls every now and then." (Abdellah)
- (35) ka-y-trini *altijd*, (..) walakin waḥed ka-y-trini *heel weinig* ..
 ASP-3-train always but someone ASP-3-train very little
 "He always trains (and he will become an acceptable soccer player), but someone who hardly ever trains .." (Hocine)
- (36) ka-n-xeyyeṭ *meestal*
 ASP-1-sew mostly
 "I sew mostly." (Fatima)
- (37) *soms* ma ka-y-ṣežb-u-ni-š
 sometimes NEG ASP-3-please-PL-1SG-NEG
 "Sometimes I don't like them [i.e. Rifians]." (Fatima)

The following adverbs of degree were found in the data:

- (38) *bijvoorbeeld, ja*, kayn-a *ongeveer* neṣṣ neṣṣ ṛ-ryafa *bijvoorbeeld*
 for example yes EXIST-F approximately half half DEF-Rifian-PL for example
 "For example, yes, there are about fifty percent Rifians for example." (Fatima)
- (39) ka-n-šuf *vooral* baramež *achter het nieuws, hier en nu, tros aktua*
 ASP-1-see especially programme-PL [title] [title] [title]
en dat soort ding-en
 and that kind thing-PL
 "I watch especially the programmes *Achter het nieuws, Hier en nu, TROS Aktua* [current affairs programmes] and that kind of stuff." (Hocine)
- (40) *min of meer* t-ḥafed ṣla l-cultuur dyal-ek
 less or more 2-preserve on DEF-culture of-2SG
 "You preserve your culture more or less." (Hocine)

7.2.3 Manner adverbs

Most of the embedded adverbs fall into this category. A large number of the embedded manner adverbs are adverbially used adjectives. In Dutch many adjectives can occur as adverbs without any morphologically marked derivation. Manner adverbs

are often modified by a degree adverb, in particular the word *heel* ‘very’. The distribution of embedded manner adverbs among the respondents is as follows: Hocine 7 tokens, Samir 6, Fatima 3, Najib and Warda 1. Bearing in mind that Samir participated in nine recording sessions and the other respondents in just one or two, it is evident that Dutch manner adverbs are relatively frequent in Hocine’s CS variety.

First consider the adverbially used adjectives in (41)–(44). Note the comparative form *helderder* ‘more clearly’ in (42), which is modified by the degree adverb *veel* ‘much’.

- (41) šuf waš der-t-ha *goed!*
 look·IMP Q do-1SG-3F properly
 ‘Look whether I did it properly!’ (Samir)
- (42) ka-t-ži-li-ya bħal ka-nšuf l-ħwayež, ka-n-šuf-ħüm *veel helder-der*
 ASP-3F-come-to-1SG like ASP-see DEF-thing·PL ASP-1-see-3PL much clear-COMPAR
 ‘It occurs to me as if I see the things, I see them much more clearly.’
 (Samir)
- (43) ka-y-kun-u ħend-ek *onbewust*
 ASP-3-be-PL at-2SG unconsciously
 ‘You will have them unconsciously.’ (Najib)
- (44) ka-t-ži *volkskrant* yumiyen ka-ne-qra-ha, ma ka-ne-qra-ha-š *precies*,
 ASP-3F-come Volkskrant daily ASP-1-read-3F NEG ASP-1-read-3F precisely

maar ka-ne-qra gewoon globaal
 but ASP-1-read just roughly
 ‘I get *de Volkskrant* [a newspaper] every day, I read it, I don’t read it precisely, but I read it just roughly.’ (Hocine)

Besides the above examples there are three other ones not cited here: *heel rustig* ‘very steadily’ (Samir) and *heel goed* ‘very well’ (both Hocine and Fatima). These examples of adverbially used adjectives contrast with attributive adjectives, which are hardly inserted at all. The next examples concern Dutch words that only occur as adverbs.

- (45) ġadi eh *vanzelf* ġadi ye-t-ġeyyer *gedacht-en*⁶ dyal-u
 FUT er automatically FUT 3-MP-change idea-PL of-3M
 ‘His ideas will change automatically.’ (Samir)

⁶ Samir makes an agreement error here: the inanimate plural noun *gedachten* is the Subject of the masculine singular verb *yetġeyyer*. With inanimate plurals, either plural or feminine singular agreement is possible in MA.

- (46) waš ža-w toevallig wella ɖerɓ-u-l-kūm t-tilifun?
 Q come-PL by chance or hit-PL-to-2PL DEF-telephone
 “Did they come by chance or had they called you?” (Samir)
- (47) u had d-dirasa lli ka-ne-tbeɣ-ha ka-n-dir-ha heel graag
 and DEM DEF-study REL ASP-1-follow-3F ASP-1-do-3F very gladly
 “And this study that I’m following, I really like doing it. (Samir)
- (48) ka-te-žmeɣ hadak š-ši, ka-t-dir-ha netjes?
 ASP-2-collect DEM DEF-stuff ASP-2-do-3F decently
 “You collect this stuff, you do it decently?” (Hocine)

The adverb *samen* “together” occurs in both Warda’s and Hocine’s contributions:

- (49) la, ka-y-xerž-u samen
 NEG ASP-3-go-out-PL together
 “No, they go out together.” (Warda)
- (50) nee, n-dir-u-h samen, maši t-dir-u nta, n-dir-u-h samen, snap je,
 no 1-do-PL-3M together NEG 2-do-3M 2M 1-do-PL-3M together understand you
 n-dir-u wahed samen b l-Šerbiya
 1-do-PL one together with DEF-Arabic
 “No, we’ll do it together. It’s not you who’ll do it, we’ll do it together, you understand? We’ll do one [presentation] together in Arabic.” (Hocine)

7.2.4 Modal and conjunctive adverbs

So-called modal adverbs express the speaker’s subjective evaluation of what she says and her attitude toward the interlocutor(s). These two uses cannot always be sharply distinguished. Several adverbs for instance emphasise the speaker’s determination about her statement, or, conversely, tone down the statement. The speaker’s determination may depend, of course, both on the content of her utterance and on the interlocutors. The distribution of modal adverbs is very different again from that of manner adverbs: while the latter are very frequent with Hocine, as compared to the other respondents, the modal adverbs are mainly inserted by Fatima. The exact distribution of modal adverbs, with the exclusion of the extremely frequent use of *bijvoorbeeld* “for example” by Fatima, is: Fatima 7 tokens, Samir and Maryam 2, Hocine 1. While embedded manner adverbs follow the MA verb, modal adverbs tend to occur clause-initially.

Embedded adverbs that qualify a statement are *bijvoorbeeld* “for instance” (very frequent in Fatima’s CS variety, cf. also (38) above, (56) below and (9) in section 1 of this chapter), *eigenlijk* “really, actually” and *misschien* “maybe”:

- (51) hna-ya *bijvoorbeeld* ka-y-ziyyin-u-ha-l-ek
 here-EMPH for example ASP-3-decorate-PL-3F-for-2SG
 “Here for example they make things seem better for you than they really are.”
 (Fatima)
- (52) ja, huwa *eigenlijk* ra-h mġeyyer
 yes 3M actually PRES-3M change-PARTICIPLE
 “Yes, it is, it has changed, really.” (Fatima)
- (53) kber-t mġa huḷandġi-yin *eigenlijk*
 grow-up-1SG with Dutch-PL actually
 “I grew up with Dutch people, actually.” (Maryam)
- (54) xexx-hūm *eigenlijk* xexx-hūm ye-qra-w *islamologie*
 must-3PL really must-3PL 3-study-PL Islamic studies
 “They should, they should study Islamic studies, really.” (Maryam)

The next example contains both the affirmative *wel* and the qualifying adverb *misschien* “maybe”.

- (55) hadak *wel* muškil, *misschien*
 DEM AFFIRM problem maybe
 “This IS a problem, maybe.” (Samir)

Adverbs that strengthen the statement are *wel* in Samir’s example (55) above as well as *zeker* “definitely”, *natuurlijk* “of course” and *echt* “really”, which occur in Fatima’s data:

- (56) ka-te-žleb ʃibad llaḥ, *bijvoorbeeld*, ja *natuurlijk* ka-te-žleb
 ASP-3F-attract servant-PL God for instance yes of course ASP-3F-attract

 ʃibad llaḥ
 servant-PL God
 “It attracts people, for instance, yes of course it attracts people.” (Fatima)
- (57) ana, ana, *zeker* xexx-ni y-welli meslem
 1SG 1SG absolutely must-1SG 3-become Muslim
 “As for me, I would definitely require that he become a Muslim.”
 (Fatima)
- (58) ila kan-et l-mra *echt* xexx-ha t-dir dak š-ši lli bġa-t, (..)
 if be-3F DEF-woman really must-3F 3F-do DEM DEF-thing REL want-3F
 “If women really have to do what they want (they should study the Koran and see whether they should or should not go out).” (Fatima)

- (59) *soms ka-ye-t-qedder te-qder te-t-zewwež b wahed lli echt*
 sometimes ASP-3-MP-make·possible 2-can 2-MP-marry with someone REL really

ka-ye-bği-k u f nefis l-weqt baği-k řla meřlaht-u
 ASP-3-love-2SG and in same time want·PART-2SG for benefit-3M
 “Sometimes it is possible, you can marry someone who really loves you and at the same time wants you for his own benefit.” (Fatima)

The adverb *gewoon* “just, no more than” occurs twice as an embedded form in Fatima’s speech, see both examples below. Hocine also embeds it once, as a modifier of another embedded adverb, see *gewoon globaal* “just roughly” in the second line of (44).

- (60) *ka-ne-qra gewoon bař n-welli parkeerwachter*
 ASP-1-study just COMP 1-become traffic·warden
 “I just study in order to become a traffic warden.” (Fatima)

- (61) *řa-te-bqa wahed l-řalaqa binat-na walakin ř-řadaqa*
 FUT-3F-remain INDEF DEF-relationship between-1PL but DEF-friendship
gewoon, maři eh..
 just NEG er
 “There will still be a relationship between us, but just friendship, not er..” (Fatima)

Concerning conjunctive adverbs, I only found *toch* “still, yet” used by Fatima. It signals a contrast with a preceding sequence in the discourse. In the quoted passage Fatima relates that marriage inevitably involves leaving your parents and, despite this pain, you will one day marry and you can still see your parents from time to time.

- (62) *řa-h toch řa-t-welli řend l-walid toch řa-te-mři te-t-zewwež*
 PRES-3M still FUT-2-come·back at DEF-father still FUT-2-leave 2-MP-marry
 “But you will still come back to your father. And yet you will leave and marry.” (Fatima)

7.3 Summary

Relatively few of the embedded PPs are identified as complements. This is in line with the tendency for embedded Dutch verbs and their complements to be realised in the same language, as observed in the preceding chapter. In section 1.1 of the present chapter it was argued that at least two out of the six attested complement PPs reveal a Dutch subcategorisation pattern, because they constitute an idiomatic expression together with the Dutch equivalent of the MA verb that surfaced in the

utterance. When we compare embedded PPs to embedded NPs in MA/Dutch, two differences between these categories attract our attention. Firstly, PPs seem to be more like the average PP in Dutch, while embedded NPs show a bias towards particular types, viz. NPs determined by a quantifier or a possessive pronoun. In other words, there appears to be a constraint on the insertion of Dutch articles which applies to embedded NPs but not (or less so) to PPs. Secondly, embedded PPs turn out to be more evenly distributed among the respondents. However, the total number of instances is rather low and, as in the case of embedded NPs, Hocine contributes relatively many tokens.

With embedded adverbs, the distribution among the respondents turned out to be strikingly unbalanced. Two speakers account for most of the case; Hocine, who inserts primarily manner adverbs, and Fatima, inserting mainly modal adverbs. With the exception of adverbially used nouns, all embedded adverbs are EL constituents, or else they combine with another Dutch adverb to form a larger adverbial constituent. They are always clausal or sentence adverbs; they do not modify MA adjectives or adverbs. In this respect they resemble embedded adjectives that do occur ‘on their own’ as predicates (or indeed as adverbs) but seldom form part of an Arabic NP or AdjP.

While embedded manner adverbs follow the MA verb, modal adverbs often precede the verb and occur clause-initially. This probably reflects the fact that manner adverbs modify the verb, while modal adverbs modify the entire clause or sentence. The word order of the other embedded adverbs is more variable. In view of the observation made concerning CS with certain other language pairs that embedded sentence adverbs tend to retain EL word order characteristics (cf. Ch. 3) it would be interesting to investigate whether the word order of embedded adverbs is best explained by MA or by Dutch syntax. However, word order properties of various adverbs are rather complicated, and this question cannot be readily answered for MA/Dutch at this point.

Chapter 8

Dutch Clauses and Discourse Markers in MA

This chapter, like the one preceding, deals with two topics. First we will discuss Dutch clauses that are embedded in a MA matrix clause. After that, we leave the domain of insertional CS, which is the focus of the present description. In order to give a more complete picture of MA/Dutch CS, in section 2 I will present some examples of Dutch discourse markers used in MA contexts. Section 3 provides a summary.

8.1 Embedded clauses

Nearly all Dutch clauses that occur as embedded constituents in a MA clause can be divided into one of four classes, according to their function in the MA clause. These are relative clauses, conditional clauses, clauses that function as the complement of a MA verb, and clauses that occupy an adjunct position. As a special type of complementation, reported speech will be discussed separately. One instance is not covered by this classification, namely the nominalised clause which modifies an (embedded) noun as the possessor in a possessive construction. This instance was quoted in Chapter 5, and is repeated here for convenience.

- (1) u ʃad te-bqa waḥed _-probleem dyał 'je moet zoveel diploma-s
and still 3F-remain INDEF problem of you must so-many diploma-PL

hebben om dat te doen'
have for that to do
“And there'll always remain a problem of 'you need so many diplomas to do that'.” (Fatima)

Setting aside reported speech and the above example, all embedded clauses are subordinate clauses and are marked by the distinctive verb-final word order in Dutch.

8.1.1 Relative clauses

The insertion of relative clauses in MA/Dutch CS deserves special attention because both languages have a different type of relative clause. Dutch makes use of so-called relative pronouns. The relative pronoun not only marks the relative clause as such, but also represents the head noun in the relative clause. When the referent of the head

“Are there any things over there in Morocco that attract you or something, let’s say that please you?” (Samir)

Example (5) is similar to (4), but this time the MA relative clause marker *lli* is also present. A further instance was produced by Abdellah, and here too, the Dutch relative clause is preceded by the MA particle *lli*. Note that in both cases the speaker interrupts his flow after *lli*, as signalled by the MA particle *zeɣma*. This particle, glossed here as epistemic, has several modal uses but also marks hesitation (Caubet, n.d. [1995]).

- (5) hadak l-weqt lli zeɣma eh *die vrijkomt* hadak t-qedd t-dir
 DEM DEF-time REL EPIST er that becomes-available DEM 2-can 2-do
- gebruik-en* baš te-ɣti-hūm zeɣma l-mawadd ž-ždid-a
 use-INF COMP 2-give-3PL EPIST DEF-subject-pl DEF-new-F
- “This time that er, that becomes available, this you can use to teach them new subjects.” (Samir)

- (6) wa l-luġa lli zeɣma *waar ik me goed in kan uitdrukken* u ..
 and DEF-language REL EPIST where I myself well in can express and
- “And the language that let’s say, that I can express myself well in and ..”
 (Abdellah)

Abdellah’s relative clause is similar to Samir’s in (3): the head noun is non-human and recurs in the relative clause as the complement of a preposition, hence the pronominal form *waar*. However, in this case the word order is different: the preposition is separated from the relative pronoun and occurs at the right-hand side of the clause, preceding the verbs. Both word orders are possible in Dutch.

The double marking of the relative clause by MA *lli* and a Dutch relative pronoun can be explained as the result of rephrasing the relative clause in Dutch after it was started in MA. Such a pattern is also found with MA discourse markers and topical constituents that precede Dutch clauses or, in the case of nominal and adverbial constituents, are incorporated as the first constituent in a Dutch clause. This will be demonstrated in the chapters on Dutch as the ML.

8.1.2 Conditional clauses

Embedded Dutch conditional clauses are found three times with Samir and once in both Hocine’s and Maryam’s data. The conditional clauses are marked by the conjunction *als* and, in Maryam’s example, *wanneer* “when”. Hocine fails to apply Dutch subordinate clause word order (the finite verb *hebt* should occur after *niks*).

- (7) *als jullie eerlijk zijn, ma ġadi-š n-ḍerb-u t-tilifun l eh eh eh*
 if you·PL honest are NEG FUT-NEG 1-hit DEF-telephone to er er er
 l-walid dyal-kūm
 DEF-parent·PL of-2PL
 “If you are honest, we will not phone er er your parents.” (Samir)
- (8) *als je een meisje ontmoet, eh ġadi t-ḥebb-ha?*
 if you a girl meet er FUT 2-love-3F
 “If you meet a girl, will you love her?” (Samir)
- (9) *als je hebt niks te doen, ṣlaš ma te-mši mṣa-h*
 if you have nothing to do why NEG 2-go with-3M
 “If you’ve got nothing to do, why don’t you go with him?” (Hocine)
- (10) *wanneer je met een marokkaanse vrouw praat, kūll-hūm ka-y-ḥasb-u*
 when you with a Moroccan woman talk all-3PL ASP-3-consider
 b ṛus-hūm naqeš ṣla ṛ-ṛažel
 with head·PL-3PL inferior to DEF-man
 “Whenever you talk to a Moroccan woman, all of them consider themselves to be inferior to men.” (Maryam)

8.1.3 Embedded Dutch complement clauses

In Chapter 6, section 4.1, we have seen three Dutch clauses as the complements of embedded Dutch verbs, plus another one in section 6. These three were embedded yes/no questions. The Dutch subordinate clauses that occur as complements of MA verbs are also for the most part interrogative clauses. Apart from those discussed in Chapter 6, I found six subordinate clause complements (3 by Samir, and one each by Fatima, Younes and Jamal). A much more frequently occurring type of embedded complements is found in so-called reported speech. In reported speech, a Dutch sequence, often a full sentence or more, is the complement of a MA verb like “to say”. I will return to reported speech later in 1.4. First, let us consider other types of complement clauses, starting with the embedded questions:

- (11) *w-ana-ya ma ka-ne-ṣref-š wa.. eh tot wanneer¹ ik moet gaan*
 and-1SG-EMPH NEG ASP-1-know-NEG wh.. er till when I must go
 “I don’t know er how far I should go [in this behaviour].” (Samir)

¹ The expected question word here would actually be *waar* “where” instead of *wanneer* “when”.

- (12) mnin n-šewwɾ-u ma ʃarf-in-š zeʃma l- eh eh *hoe de*
 when 1-recreate(?) -PL NEG know*PART-PL-NEG EPIST DEF-er how the

maatschappij in elkaar zit
 society in each·other sits

“When we walk around [in Morocco] we’re not aware of, well, what the society is really like.” (Younes)

- (13) hadak š-ši ma ne-ʃref-t² ana-ya, *wat hij wil worden*
 DEM DEF-thing NEG 1-know-1SG 1SG-EMPH what he wants become
 “This I don’t know, what he wants to become.” (Jamal)

Hocine’s embedded complement clause is not a question. It is part of a particularly curious and complicated bilingual utterance: the governing verb is itself in Dutch but the inflection is MA (this was discussed as a minor pattern of verb insertion, Ch. 6, section 6), the complementizer is the MA *belli*, and the complement clause is Dutch again. This time Hocine correctly applies Dutch subordinate clause word order. I repeat the example here.

- (14) ʃref-ti ka-t-bewijs belli ‘t goed is, *snap je?*
 know-2SG ASP-2-prove COMP it right is understand you
 “You see, you prove that it’s right, you understand?” (Hocine)

Finally, Samir inserts two Dutch clauses that are not syntactically marked as such: there is no overt complementizer, and the Dutch clauses lack the subordinate clause word order.

- (15) gūl-t ana-ya ġadi n-šuf eh *misschien heb, ben ik fout of is hij eh*
 say-1SG 1SG-EMPH FUT 1-see er maybe have am I wrong or is he er

fout, weet ik niet
 wrong know I not

“I said [to myself] I will see er maybe I’m, I’m wrong or he’s er wrong, I don’t know.” (Samir)

- (16) f l-qurʔan ma ka-t-bin-š laylat l-qadr *is op de zevenentwintigste*
 in DEF-Koran NEG ASP-3F-appear-NEG Laylat l-Qadr is on the twentyseventh

dag dyaḷ ramḍan, of wel?
 day of Ramadan of AFFIRM

² On this peculiar verb form see note 12 on p. 253.

“In the Koran it is not said [that] Laylat l-Qadr is on the 27th of Ramadan, or is it?”³ (Samir)

In the latter example I consider *laylat l-qadr* and *dyal ramḍan* to be MA constituents in the Dutch copula clause.

8.1.4 Reported speech

A well-known stylistic function of CS is its use in narratives to distinguish between the main text and the dialogues of the characters. This is also a recurrent feature in the Nijmegen corpus. I have not studied the phenomenon in detail, because although the stretch of reported speech is formally a complement of a verb that denotes a speech act, the regularities and constraints dictating its use are not of a syntactic nature. Indeed, any sequence of words in any language can be the complement of a verb like “to say”. For this reason, the study of CS for reported speech is better handled in terms of discourse analysis and text grammar. In the present description of the MA/Dutch corpus, I will just give one example to demonstrate how CS for dialogues works in a narrative text. In MA/Dutch CS the dialogues are usually in Dutch while the connecting text is in MA. As the narrative is often set in the Netherlands, the choice of Dutch for the dialogues will often mirror the language actually used in the related event, although this relationship does not always apply. Language choice does have an important communicative function in the following example:

(17) narrative by Samir

gal-et-l-u eh ‘*ik ben wel een beetje verliefd op jou*’
say-3F-to-3M er I am AFFIRM a bit in·love with you

iwa u men temmak gal-l-ha, skūt šwiya ma gal-l-ha ta ši
well and from there say-to-3F keep·silent a·little NEG say-to-3F even INDEF

ḥaža u fatali gal-et-l-u ‘*zeg je niks nou?*’
thing and then say-3F-to-3M say you nothing now

gal-et-l-u ‘*ik ben verliefd op jou*’
say-3F-to-3M I am in·love with you

³ In the Koran (surah 97) it is said that the revelation of the Koran took place in the night called *Laylat l-Qadr*. This is traditionally believed to be on the 27th of Ramadan.

gal-l-ha 'ik niet op you'
say-to-3F I not with you

gal-et-l-u 'kifaš? omdat ik zeker marokkaans meisje ben hè?
say-3F-to-3M how because I certainly Moroccan girl am QTAG

daarom vind je mij niet leuk?
for.this find you me not nice

gal-l-ha gal-l-ha 'la, omdat je geen marokkaans meisje bent!
say-to-3 say-to-3F NEG because you no Moroccan girl are

(A Moroccan girl meets a boy in a disco. She thinks he is a Dutch guy, but then he turns out to be Moroccan too.)

"She said to him: 'I er I'm a bit in love with you, actually.'

Well, and then he said to her, he kept silent for a moment, he didn't say anything to her. And then she said: 'Aren't you going to say anything now?'

She said: 'I'm in love with you.'

He said to her: 'I'm not in love with you.'

She said: 'Why? It's because I'm a Moroccan girl, isn't it? That's why you don't fancy me?'

He said to her, he said: 'No, it's because you're not a Moroccan girl.'

(Moral: a real Moroccan girl does not pick up boys in the disco.)

In this story it is particularly functional that the dialogues are in Dutch. Otherwise, how can the girl mistake a Moroccan for a Dutchman when they are having a conversation in Arabic?

8.1.5 Adverbial clauses

Samir embeds a number of clauses that function as an adverbial adjunct in the MA clause. I have five examples of this. Four of these are reproduced here, and see further ex. (47) in Ch. 6. These are all well-formed subordinate clauses in Dutch. (18) and (19) show a qualifying adjunct; ex. (47) in Ch. 6 and (20) below can be classified as manner adjuncts.

- (18) voor zover ik jou ken, l-maʃiʃa temmak ma ka-t-ži-l-ek-š gaʃ
for so-far I you know DEF-life there NEG ASP-3F-come-to-2SG-NEG at.all
"As far as I know you, you won't like the way of life over there at all." (Samir)

- (19) hiya voor zover ik weet eh ehm ma ʃend-ha-š ouder-s
3F for so-far I know er er NEG at-3F-NEG parent-PL
"As far as I know, she er er has no parents." (Samir)

- (20) zeŋma hoe ik ben, ne-hdeɣ mŋa-hŋm hakda
 EPIST how I am 1-talk with-3PL like-this
 “The way I am I talk with them this way.” (Samir)

The Dutch clause introduced by *om* “in order to” in (21) expresses the purpose of *gebruiken* “to use”. The complementizer *om* requires an infinitive verb complement which is marked by the particle *te*. In MA such a ‘purposive’ clause is introduced by the complementizer *baš* followed by the subjunctive mood, like in the second line of example (5) above. The interchangeability of these MA and Dutch clause types has been noted several times in this corpus description; see further Chapter 11 on embedded MA clauses.⁴

- (21) ntaŋ aš luğa - b d-dariža - lli ka-te-ŋžeb-l-ek baš eh eh tdir-ha
 of what language with DEF-MA REL ASP-3F-please-to-2SG COMP er er 2-do-3F
gebruiken om jouw gevoelens uit te drukk-en, ntaŋ aš luğa?
 use-INF COMP your feelings express to express of what language
 “What language, [tell me] in Moroccan Arabic, do you like to use to express your feelings, what language?” (Samir)

8.2 Discourse markers

As we discussed in Chapter 3, various word classes and expressions are used for discourse marking. Discourse marking in turn covers a wide range of functions that can be broadly divided into two types: the expression of the speaker’s attitudes, and the ordering of the sequential structure of the discourse. In the preceding chapter we already examined Dutch prepositional phrases and adverbs that serve either of these functions while being constituents within in a MA clause. In addition to these, there are various Dutch words and phrases that serve discourse marking functions in a MA context but are not easily analysed as embedded elements. This concerns in particular the markers that occur clause-initially and clause-finally. (See the discussion in Chapter 3.) For this reason, the study of discourse marking falls partly outside the competence of the Monolingual Structure Approach, which is aimed at syntactic and morphological analysis.

The respondents use both Dutch discourse markers in MA contexts and MA markers in Dutch contexts. It is very well possible, or even probable, that speakers

⁴ MA ‘*baš*-clauses’ occur as complements of embedded Dutch nouns (Ch. 5, section 1.9), as complements of embedded Dutch verbs (Ch. 6, section 4.2), as Subject in a Dutch matrix clause (Ch. 9, section 3.1.3) and as complement of a noun in a Dutch clause (Ch. 9, section 3.1.4).

prefer discourse markers from either language for certain functions. It can be observed, for instance, that MA conjunctions recurrently conjoin Dutch clauses (see Ch. 9, section 3.2) while the reverse, Dutch conjunctions conjoining MA clauses, occurs rarely. We can also expect considerable idiosyncratic differences given the diversity of linguistic and cultural backgrounds amongst the respondents of the Nijmegen corpus. A thorough investigation of this interesting topic is, unfortunately, beyond the scope of the present study. I will content myself with giving some illustrative examples.

8.2.1 Conjunctions

There are few instances of Dutch conjunctions preceding MA clauses. MA conjunctions more often precede Dutch clauses, as we will see in Chapter 11. However, some examples can be cited: the adversative *maar* “but” in (22) marks an argumentation that is phrased in MA; another example is found in the third line of (27) in Chapter 7. The Dutch coordinate conjunction *of* “or” conjoins two MA clauses in (23), and in (65) above *of* joins a MA subjunctive clause to a preceding Dutch infinitive.

- (22) *ik weet niet of jij het goed gelezen hebt, maar* huwa gal f l-maqal
I know not whether you it well read have but 3M say in DEF-article

dya-l-u l-meslem-in qell men eh l-axŭr
of-3M DEF-Muslim-PL less than er DEF-other

“I don’t know whether you’ve read it well, but he said in his article [that] Muslims are less than er what’s it ..” (Samir)

- (23) *ila ma y-dir-š mša-k mezyan natuurlijk, of y-xenzer fi-k ofzo*
if NEG 3-do-NEG with-2SG well of course or 3-scowl at-2SG or something
“If he doesn’t treat you well, of course, or scowls at you or something ..”
(Fatima)

8.2.2 *ja* and *nee*

Fatima and Zineb repeatedly use Dutch *ja* “yes” and *nee* “no” at the beginning of a turn, as a means to take the floor. In addition *ja* and *nee* convey a concessive meaning; “yes/no, you are right (but)”.

- (24) (Zineb has stated, in Arabic, that the faculty students have a rather high intellectual level. Because of this you don’t notice their discriminatory attitudes. Then Samir replies in Dutch that this is actually disappointing, and Zineb retakes the floor:)

ja, ja maši küll-ši, ma ka-nehḍeṛ-š ʃla küll-ši
 yes, yes NEG everyone NEG ASP-1-talk-NEG about everyone
 “Yes, sure, not everyone, I’m not talking about everyone.” (Zineb)

- (25) ka-n-heḍṛ-u ʃla ž-génération lli qṛa-t, ʔamma duk lli ma
 ASP-1-talk-PL about DEF-generation(Fr) REL study-3F as-for DEM REL NEG

qṛa-w-š, l-msaken, *ja*, ila žṛa ʃli-ha ṛ-ṛažel, ka-t-gul-l-ek ḡadi
 study-PL-š DEF-poor-PL yes if attack on-3F DEF-man ASP-3F-say-to-2SG FUT

ne-ṛžeʃ ʃend walid-i-ya

1-return at parent-PL-1SG

“We’re talking now about the generation (of women) that studied. As for those who didn’t study, poor things, well, yes, if her husband attacks her, she’ll say ‘I’ll go back to my parents.’ ” (Zineb)

- (26) (Samir is surprised that Fatima is preparing to become a traffic warden. Then Fatima asks him why he thinks this is strange.)

S la xaṭeṛ nti-ya bḡi-ti te-ṛžṛi l l-meḡrib.
 because-2F-EMPH want-2SG 2-return-F to DEF-Morocco
 “Because you wanted to go back to Morocco.”

F *ja*, ana ta t-ʔakked-t bi-ʔanna ma ne-qḍeṛ-š ne-ṛžeʃ l
 yes 1SG until MP-assure-1SG COMP NEG 1-can-NEG 1-return to

l-meḡrib, ʃad xṭar-t had eh ..
 DEF-Morocco then choose-1SG DEM er

“Yes, it was only after I became convinced that I can’t go back to Morocco that I chose this er [project].”

(Fatima & Samir)

8.2.3 The question tag *hè*

Several respondents make ample use of the Dutch question tag *hè* [hɛ] (with rising intonation). This particle, which invites the addressee to signal that he is still paying attention to the speaker, occurs at the end of a constituent, a clause, or an utterance.

The following narrative text by Younes comes from the conversation in which all four Hamadi siblings and their mother participated. At the beginning of his story, Younes uses *hè* repeatedly to assure himself of the attention of his audience, and the particle recurs once more towards the end of the story.

- (27) ne-ʃref waħed l-weld eh y-dir *karate h  *, l-kara ti *h  *,
 1-know INDEF DEF-guy er 3-do karate QTAG DEF-karate QTAG
- u ʃend-u eh ʃend-u ... la, ʃend-u *zwart-e band h  *,
 and at-3M er at-3M NEG at-M black-AGR belt QTAG
- huwa   a-h kbir,   a-h ʃ  rin sna, n-ʃer f-u ana, qari f *paulus*
 3M PRES-M old PRES-M twenty year 1-know-3M 1SG study PART at Paulus
- lyceum*, qari ze ma, *VWO* [feweo] y-dir, u ze ma    a-h ila
 Lyceum study PART EPIST VWO 3-do and EPIST father-M if
- ye-fhem   li-h y-    b-u, y    b-u   etta y-  eyy  -u f l-er  .
 3-understand on-3M 3-hit-3M 3-hit-3M until 3-make fall-3M at DEF-ground
-     al men xet  ra y-    b-u u     a-h ma ma y-dir-l-u walu *h  *,
 how many of time 3-hit-3M and father-M NEG NEG 3-do-to-3M anything Q T A G
- y-xaf menn-eh. u   ad saken f         
 3-be afraid of-3M and still live PART in DEF-house
- ‘‘I know a guy who does karate, you know, karate, right? And he has, he has
 .. No, he has the black belt, right? And he’s old, he’s 20 years old. I know him,
 he goes to school at the Paulus Lyceum. He’s a good student, actually. He does
 VWO [pre-university education]. And when he’s obstinate he hits his father.
 He hits him until he knocks him down. He hits him so many times, and his
 father can’t do anything to him, you see? He’s afraid of him. And he still lives
 with his parents.’’ (Younes)

8.3 Summary

The insertion of entire Dutch subordinate clauses is not frequent in the corpus, and is confined to the respondents who used both languages for entire sentences and also engaged in intersentential CS during the recordings. The greater number of the examples were in fact produced by Samir. The embedding of Dutch clauses in MA main clauses has syntactic consequences that are adequately described in terms of ML structure and EL constituents. The embedded Dutch clauses are marked as subordinate clauses by means of word order. Embedded Dutch complement clauses in particular clearly occupy the syntactic position of a constituent in the MA matrix clause. The embedded clauses appear to be well-formed EL constituents, and they are more similar to their counterparts in monolingual Dutch than to embedded NPs.

Discourse marking in bilingual text takes place in both languages. This study will not investigate the use of discourse markers in detail. However, some conspicuous patterns were pointed out in this chapter, notably the use of the Dutch words for ‘‘yes’’

and “no” as markers of turn taking and as markers of concession “yes, but..” and the use of the Dutch question tag *hè* in MA discourse. The importance of these observations lies in the fact that they demonstrate another aspect of the modularity of speech production: these aspects of discourse marking constitute a module that is independent from clause syntax, in the sense that either one of these can evolve in a different language.

Chapter 9

MA Insertions in Dutch

Because Moroccan Arabic insertions in Dutch matrices are less numerous than Dutch insertions in Moroccan Arabic matrices, the former will all be discussed in the present chapter. Absolutely no MA verbs were found to be embedded in Dutch matrices. Apart from this, the organisation of the sections in this chapter parallels the description of Dutch insertions in MA in the preceding Chapters 5 to 8. Section 1 is concerned with the nominal and adjectival categories. Section 2 addresses the insertion of MA prepositional and adverbial constituents, while the embedding of MA clauses and the use of MA discourse markers is treated in sections 3.1 and 3.2, respectively. A summary will conclude each of sections 1, 2 and 3. As will become clear in the course of this chapter, the insertion of Dutch and MA elements does not only differ quantitatively, but there are also important qualitative differences.

The main findings concerning the Nijmegen corpus will be recapitulated in section 4 of this chapter. First I will point out the major individual differences between the respondents with respect to their codeswitching behaviour. Finally, an overview table will be presented which summarises all insertion types attested in both MA and Dutch matrices.

9.1 MA nouns, NPs and adjectives in Dutch matrices

The insertion of MA nouns, adjectives, and nominal constituents in Dutch matrix structures yields a very different picture from what we have seen in Chapter 5. In the first place MA insertions are far less numerous. For this reason it is not possible to investigate the distribution of MA nouns in Dutch matrices in detail, such as was done in Chapter 5 for embedded Dutch nouns. Therefore we cannot establish to what extent the distribution of MA nouns parallels the distribution of Dutch (ML) nouns. Yet with respect to the available examples we can observe some striking qualitative differences between Dutch and MA embedded elements. MA insertions turn out to be more readily associated with certain identifiable factors, namely culturally specific vocabulary, repetition in discourse, and topicality.

9.1.1 Insertion of nouns and adjectives

There are only a handful of MA nouns in Dutch clause, all of which will be cited here. The occurrence of embedded nouns can be traced back to two factors that are both highly specific, albeit of a very different nature. The majority of the insertions

consists of terms that refer to typically Moroccan or Islamic culture and society. These are all nouns or fixed nominal expressions. Embedded MA content words that do not belong to this category are even rarer, and can be related to a discourse phenomenon called topic continuity. These are MA nouns and adjectives that recur in a Dutch structure after they have been used in a MA matrix in the immediately preceding discourse.

9.1.1.1 Moroccan and Islamic terminology

One could claim that many of the Dutch insertions in MA also refer to concepts related to Dutch society and the speaker's experiences in the Netherlands. So Dutch *dorp* "village" in the first example of Chapter 5 refers indeed to a particular village in the Netherlands, and various terms that refer to the Dutch educational system feature in conversations in which the respondents discuss their studies, which also take place in the Netherlands. However, these terms generally have translation equivalents in MA and the respondents often know the MA words for these concepts, although it can be argued that the MA terms never have quite the same associations as the Dutch ones. Yet the MA nouns in Dutch are on average much more culturally specific and less translatable than the Dutch nouns discussed in Chapter 5. I found eight instances (seven tokens) all of which are cited hereafter.

The term *ḥizb* in (1) for instance refers to a portion of the Koran. For educational and devotional practices the text of the Koran is divided into 60 portions of equal length called *ḥizb*, plural *ṭaḥzaab* in Classical Arabic. In Morocco this partitioning is used for learning the Koran by rote: one memorizes one *ḥizb* after the other, until one knows the entire Koran by heart. The *ṭaḥzaab* are known by the word they begin with. *sabbih* "praise" (imperative) is the first word of the 60th *ḥizb*, so in (1) Mimoun actually claims that he succeeded in memorising the entire Koran, an achievement few Moroccans can pride themselves on.

- (1) *ik had ḥizb sabbih helemaal geleerd*
 I had hizb "praise" entirely learnt
 "I had memorised all of hizb sabbih." (Mimoun)

Other examples of Islamic terminology are *laylat l-qadr* and *ṛamḍan* cited in Chapter 8, ex. (16), partially repeated here in (2). *laylat l-qadr* and *ḥizb sabbih* are formally nominal constituents, but they are set phrases borrowed from Classical Arabic. In the Dutch finite clause they are treated as proper names. *dyaḷ ṛamḍan* in (2) is obviously a possessive PP and not a set phrase.

- (2) *laylat l-qadr is op de zevenentwintigste dag dyaḷ ṛamḍan*
 Laylat l-Qadr is on the twenty-seventh day of Ramadan
 "[The Koran does not say that] Laylat l-Qadr is on the 27th of Ramadan."
 (Samir)

In (3) Samir pronounces the name of the Prophet and Messenger of God, Mohammed, in Arabic, followed by the eulogy in Classical Arabic that conventionally follows when a Muslim mentions the prophet's name, here loosely translated as "God's peace be upon him". Also in (3), Samir uses the MA word for mosque, although the matrix language of the clause is Dutch. A determiner is missing before *žameſ* "mosque", unless the initial consonant is interpreted as a geminate; in that case *ž-žameſ* is an embedded NP. Then, in (4), the MA word *ḥadit* again refers to a very specific Islamic concept, namely the corpus of orally transmitted texts relating the deeds and utterances of the prophet and his companions.

- (3) *muḥammed ʃalla ʔlahu ſlay-hi wa sallam was es 'n keertje in* __
 Mohamed 'God's peace be upon him'(ClAr) was once a time in [DEF]
žameſ, toen kreeg hij ineens een eh openbaring
 mosque then received he suddenly a er revelation
 "One day Mohamed, God's peace be upon him, was in the mosque, and then he suddenly got a revelation, (..)" (Samir)

- (4) *volgens de ḥadit, ik weet niet hoe dat in het echt is gelopen,*
 according to the Tradition I know not how that in the real is walked
verlopen
 occurred
 "[This is] according to the Prophetic Tradition, I don't know how this happened, occurred in reality." (Samir)

In the following example MA *siyasa* "politics" refers specifically to Moroccan politics. In Morocco people regard politics and anything to do with local or national government as a dangerous issue from which one should keep a safe distance. Example (5) is extracted from a passage where Samir argues against the employment of Moroccan trained social workers in the Netherlands. The Moroccan immigrants, he claims, do not trust them because they associate them with the Moroccan government:

- (5) *zij denken dat het iets met de siyasa te maken heeft*
 they think that it something with the politics to do has
 "They think it has something to do with (Moroccan) politics." (Samir)

The next example speaks for itself. The *jellaba* is a North African type of robe with a hood. Men and women wear different kinds of *jellabas*. If you go out wearing a *jellaba* in the Netherlands you will attract the public's attention, says Fatima.

- (6) *er is één meisje of één man die met een baard en met een žellaba,*
 there is one girl or one man who with a beard and with a jellaba

lang-e žellaba bedoel ik, ik bedoel ja, marokkaanse eh kleren,
 long-AGR jellaba mean I I mean yes Moroccan er clothes
 “(Suppose) there’s just one girl (with a headscarf) or one man with a beard
 and a jellaba, I mean a long jellaba, I mean, well, Moroccan er dress, (he’ll
 attract the people’s attention).” (Fatima)

Note that in the last three examples, the MA noun is determined by a Dutch article and unequivocally embedded in a Dutch NP constituent. I will return to this point at the end of the following section.

9.1.1.2 Embedded MA words in repetitions

A rather different category of MA content words in Dutch matrix structures results from the speakers’ tendency to reuse the same content words in an ongoing discussion of a topic. (Repetition will be discussed as an explanatory factor in Chapter 11.) MA nouns and adjectives in Dutch matrices demonstrate topic continuity as a source of CS behaviour very clearly: repetition is a prominent and recognisable factor in about half of all the instances. Characteristic of EL forms that result from repetition is that they are not necessarily specific. Indeed, this mechanism accounts for nearly all embedded nouns and adjectives which do not denote specifically Moroccan or Islamic concepts. I have five examples; three embedded MA nouns by Samir and two adjectives by Hayat and Fatima.

In (7) Samir inquires about Jamal’s housing situation. Note the MA word *bit* “room” in the first line of the cited passage; it recurs in a later turn by Samir (line 3), in the Dutch NP *nog een andere bit* “yet another room”, which is the complement of MA *šend-u* “he has”.

(7) S waš šend-ek bit waḥd-a u šafi hna-ya?
 Q at-2SG room one-F and that’s all here-EMPH
 “Do you have just one room here?”

J iyeh, hiya hadi
 yes 3F DEM
 “Yeah, that’s it.”

S (..) u huwa šend-u *slaapkamer*, en *nog een ander-e* bit, *douche*¹
 and 3M at-3M bedroom and yet a other-AGR room shower
 “(..) So he [i.e. your housemate] has a bedroom, also another room, and
 a shower..”

(Samir and Jamal)

¹ The word for “shower” could be either Dutch *douche* [duʃ] or its MA cognate *duš*.

In (8) we see that the MA word *mṛa* “woman” in the first line recurs in a Dutch matrix NP *een mṛa* “a woman” in the third line. This NP is part of a larger Dutch NP *meer hersens dan een mṛa* “more brains than a women” which is itself embedded in a MA matrix clause as the complement of *ṣend-u* “he has”. The third example is *fransawiya* “French” in (20) below. This word is mentioned a number of times in MA contexts in the preceding part of a conversation about foreign languages in Morocco. The MA noun *xūbz* “bread” in example (103) in Chapter 5 is most likely another example, cf. its discussion on p. 206.

- (8) waš f ṛ-ray dyal-kūm zeṣma ṛ-ražel u l-mṛa mextelf-in,
 Q in DEF-opinion of-2PL EPIST DEF-man and DEF-woman different-PL
- zeṣma ta f l-ṣqel? zeṣma eh ṛ-ražel ṣend-u l- eh n-gul-u *meer*
 EPIST even in DEF-mind EPIST er DEF-man at-3M DEF- er l-say-PL more
- hersens dan een mṛa*, (wella had š-ši eh *is onzin?*)
 brains than a woman or DEM DEF-thing er is nonsense
 “In your (PL) opinion, are men and women different, even mentally? That is,
 men have er let’s say more brains than women, (or is this nonsense?)” (Samir)

The next two examples concern adjectives. In these cases the repetition has an obvious rhetorical effect. *mwessxa* in (9) is used attributively; note the MA word order noun-adjective in *straat mwessxa* “dirty street”.² The Dutch PP *langs een straat mwessxa* is itself embedded in a MA matrix clause. Then *ḍ ḥifa* in (10) is part of a left-dislocated adjectival constituent that is the predicate of the Dutch copula *is* “is”. In this case the MA adjective *ḍ ḥifa* is modified by the Dutch degree adverb *zó* “that much”.

- (9) ma ta-ye-bḡi-w-š ḍyur mwessx-in *langs een straat mwessx-a*
 NEG ASP-3-want-PL-NEG house-PL dirty-PL along a street dirty-F
 “They don’t want dirty houses along a dirty street.” (Hayat)
- (10) ma ḍḥif-a-š. zó ḍḥif-a *is ze niet*
 NEG weak-F-NEG that-much weak-F is she not
 [Talking about the capacities of women as compared to men.]
 “She’s not weak. She’s not all that weak.” (Fatima)

² This NP is analysed as a Dutch NP, although it contradicts the premisses of the MSA. The EL order of the EL attributive adjective is a recurrently occurring type of counter-example (cf. Ch. 2, section 4.1.4). If *een straat mwessxa* were to be interpreted as a MA NP, the insertion of the Dutch free form article *een* would be a more serious divergence from the already established insertion patterns.

It is striking that in all three examples, (7), (8) and (9), the MA word is embedded in a Dutch matrix constituent that is itself embedded in a MA clause. These are instances of what I have called layered embedding in Chapter 2. Another embedded MA noun that also involves layered embedding, but which is less clearly the result of repetition, is found in example (3) in Chapter 7, repeated here as (11). Here the MA word *huḷanda* occurs in the Dutch PP *naar huḷanda* “to Holland”.

- (11) huma ža-w als arbeider-s naar huḷanda
 3PL come-PL as worker-PL to Holland
 “They came to Holland as workers.” (Samir)

Morphosyntactic embedding of MA nouns

The following observations pertain to all attested MA nouns in Dutch constituents, whether instances of culturally specific terminology or of content word repetition. In the case of some of the embedded MA nouns, it appears that they are assigned common gender in Dutch. This is evident from the common gender definite article *de* in (5) and (4) and the agreement suffix *-e* on the Dutch attributive adjectives in (6) and (7). Embedded plural nouns were not found, except for plural nouns prefixed by the MA article *l-*; these will be discussed in the following section.

With respect to Dutch content words in MA matrices I noted some ‘co-occurrence restrictions’: firstly, Dutch nouns are seldom modified by a MA attributive adjective; secondly, Dutch embedded adjectives mainly occur as predicates in MA copula constructions, and thirdly, neither embedded adjectives nor adverbs are ever modified by a MA adverb (with the exception of ex. (106) in Ch. 5). On the other hand MA adjectives and nouns in Dutch matrices, as far as we can judge from the small number of instances, are not subject to these kinds of co-occurrence restrictions. MA nouns are modified by Dutch attributive adjectives: *lange žellaba* “long jellaba” in (6) and *andere bit* “other room” in (7). In (9) and (10) above we saw embedded MA adjectives. *mwessxa* “dirty” in (9) modifies a Dutch noun. The predicatively used *d žifa* “weak” in (10) is itself modified by a Dutch adverb that precedes it, in accordance with ML word order. Therefore *d žifa* is part of a Dutch adjectival constituent, something which distinguishes it from the Dutch adjectives that occur as the predicate complements of MA copulas (discussed in Ch. 5, section 3.2).

In Chapter 5 it was shown that when Dutch nouns are embedded in MA matrices, the definite article, which is obligatory in the ML, was found to be missing in a large number of the cases. There is no such omission phenomenon with regard to embedded MA nouns: these are modified by Dutch articles as expected within the framework of the MSA. See *de hadit* in (4), *de siyasa* in (5), *een žellaba* in 303; only in (3) the MA word *žame ſ* “mosque” seems to lack an article. It is possible that in this case the initial consonant is actually geminate, and *ž-žame ſ* is an embedded NP (see section 2 below).

With regard to the relative order of attributive adjective and noun in these examples, we note the order Adjective-Noun, in accordance with Dutch (ML) grammar, when the adjective is Dutch as in (6) and (7), while in (9), where the adjective is MA, we find the order Noun-Adjective, in accordance with the EL. So in these three cases the language of the adjective determines the word order.

9.1.2 Insertion of nominal constituents

This section is concerned with MA constituents that occupy the position of an NP in a Dutch matrix. We find two types of embedded NPs: MA nouns that are modified by a MA determiner, and pronominal forms. Embedded MA constituents are fairly common as Topics in Dutch copula clauses. There are only a handful of examples in other contexts which will be discussed first (1.2.1); section 1.2.2 deals with the Topic NPs in copula clauses; 1.2.3 with Topic NPs in other than copula clauses. Extra-clausal Topics will be addressed in 1.2.4.

9.1.2.1 Nouns determined by an article

There are five examples of embedded nouns that are modified by the MA prefix *l-* or one of its assimilated allomorphs.³ From the viewpoint of MA these are inflected word forms and may or may not constitute full NPs. From the viewpoint of Dutch as the ML they occupy the position of a nominal constituent rather than an inflected word in Dutch. For this reason, I classify MA nouns marked by the MA definite prefix as embedded constituents. This contrasts, for example, with the treatment of Dutch plural nouns in MA matrices (Ch. 5, section 1.3), which are inflected content words in Dutch as well as in MA.

In one instance the definite prefix is not part of a geminate cluster. Consider (12) where the definite prefix *l-* occurs in the embedded form *l-kafîrîn*⁴, complement of Dutch *als* “like”.

- (12) *alle nederlanders zag ik als eh .. l-kafir-in, n-gul-u, b*
 all Dutchmen saw I like er DEF-unbeliever-PL 1-say-PL with

l-luğa dyal-na n-gul-u
 DEF-language of-1PL 1-say-PL

“I considered all Dutchmen to be er .. unbelievers, let’s say, in our language that is.” (Samir)

³ In addition I note an instance of *ši xetîr-at* [INDEF time-PL] “sometimes” which I will discuss in section 2.2 below as an embedded adverb (ex. (47)) although it is formally a NP.

⁴ This form is itself a CS form from Classical or Modern Standard Arabic. The expected MA dialect form would be *kūffar*, plural of *kafer*.

In the other four examples the presence of the MA definite prefix depends on the perception of the initial consonant as a geminate. As I mentioned in Chapter 5 (section 1.4) the perception of the distinction between simple and geminate consonants in onset position is not obvious. In the following two examples, I assume the presence of a definite prefix in *d-dariža* ‘Moroccan Arabic’ and in *ž-žnun* ‘genies’. Hence these forms are classified as embedded NP constituents. (See also the discussion of *žame* *ʃ* ~ *ž-žame* *ʃ* in (3) on p. 306.)

- (13) *(je had eh voor de middag had je les in het arabisch en na de middag dan heb je les in het frans)*

en dan komt er nog eh d-dariža f d-ḍar,
and then comes EXPLETIVE yet er DEF-Moroccan Arabic in DEF-house

(dus je zit eigenlijk met drie talen door elkaar)

“(You had er .. In the morning you had courses in [Standard] Arabic and in then the afternoon you have courses in French), and then there’s Moroccan Arabic at home, (so you’re faced with three languages at a time.)” (Abdellah)

- (14) *je moet niet ’s avonds niet naar buiten gaan, omdat je anders*
you must not at night not to outside go because you otherwise

ž-žnun tegenkomt

DEF-genie•PL meet

“You shouldn’t go outside at night, because otherwise you’ll encounter genies.” (Samir)

- (15) *nog meer dingen van r-ryafa (die ma ka-ye-šzeb-l-ek-š?)*
yet more things of DEF-Rifian•PL REL•PRON NEG ASP-3-please-to-2SG-NEG
“[Is there] anything else about Rifians that you don’t like?” (Samir)

Note that the nouns in the embedded NPs in examples (12)-(15) again refer to concepts specific to Moroccan or Islamic culture: *kafir* ‘unbeliever, i.e. non-Muslim’ and *ženn*, plural *žnun* ‘genie’ belong to the Islamic cultural heritage, *d-dariža* is the common name in Morocco for the local Arabic dialects, and *r-ryafa* ‘Rifians’ defines an ethnic group from the Moroccan perspective.⁵

⁵ In Morocco, people from the Rif area, whether Arabophones or Berberophones, are associated with backwardness and religious conservatism. Women from the urban centres in western Morocco like Zineb and Fatima reproach Rifians for their conservative attitudes in relationships between men and women. In Dutch society, when different ethnic groupings of Moroccans are distinguished at all, the distinction is made between Arabs and Berbers, with the latter group being primarily associated with a lack of formal education.

Example (16) involves repetition similar to the cases discussed in section 1.2 above. In the first line we see the embedded PP *f d-din* (which will be discussed in section 2.1), and in the third line the NP *d-din* recurs in a Dutch matrix structure, namely the PP *met d-din* “with the religion”. Like most examples cited in section 1.2, (16) involves layered embedding: the Dutch verb plus complement *mixen met d-din* “to mix with the religion” is itself embedded in a MA clause.

- (16) *(je had twijfels aan .. f d-din dyal-ek u ka-t-šuf l-walid*
 you had doubts about in DEF-religion of-2SG and ASP-2-see DEF-parent's

dyal-ek wašta y-dir-u? Šend-hūm hadak t-taqalid)
 of-2SG what 3-do-PL at-3PL DEM DEF-tradition·PL

ka-y-dir-u-ha mix-en met d-din
 ASP-3-do-PL-3F mix-INF with DEF-religion
 “You had doubts about .. about your religion, and you see your parents, what
 do they do? They have those traditions and they mix them with the religion.”
 (Samir)

With respect to EL constituents that contain function morphemes, an interesting question arises as to whether it is the language of the constituent or the language of the matrix clause (i.e. the ML) which governs the distribution of these function morphemes. The definite article in the plural forms *l-kafirin* “the unbelievers” in (12) and (if /ž/ is indeed geminate) *ž-žnun* “genies” in (14) must be attributed to MA rather than Dutch (ML) grammar. In generic expressions, MA may use the definite plural to designate a class or an unspecified number of count nouns (Caubet, 1993, II: 295-7).⁶ Dutch also has the definite plural noun marked by *de* in certain generic usages, however this is subject to complicated idiomatic restrictions (Geerts et al., 1984: 119). According to my intuition, a Dutch definite article would be impossible in (12) and (14), and Dutch would use the indefinite plural in these contexts. A Dutch definite article with a generic function would be more acceptable in (15).⁷ In the case of *d-dariža* in (13), a definite prefix is in accordance with MA grammar, and it is not ungrammatical in Dutch. In *d-din* in the third line of (16), a definite article is

⁶ Generally, only the Subject (or Topic) NP on which a predication is made is considered with respect to generic expressions. Yet the same principle accounts for the definite plural in (12): *l-kafirin* represents a class rather than specifically identifiable unbelievers (Caubet, 1993 vol. II: 295).

⁷ Perhaps this is because *r-ryafa* “the Rifians” in (15) can be interpreted as categorical, encompassing all the Rifians, whereas *l-kafirin* in (12) and *ž-žnun* in (14) are mere representatives of a class. Clearly, all Dutchmen together constitute only a part of the unbelievers, and one is unlikely to encounter the entire genie world upon leaving the house. However, there may also be syntactic constraints involved.

appropriate according to both MA and Dutch. In conclusion, on the basis of this limited evidence, my impression is that the distribution of definite articles in inserted constituents is governed by MA rather than Dutch grammar. Remember that this is not obviously apparent, since in North African Arabic/French CS the embedded French clitic definite articles are used to express Arabic grammatical features (Ch. 2 section 1.1). More examples will be considered in the next section, on embedded MA NPs in copula clauses.

9.1.2.2 Topic NPs in copula clauses

One CS pattern that occurs frequently in comparison with other instances of embedded MA nouns or NPs is the embedding of a MA constituent as the Topic and Subject of a Dutch copula clause. I have 17 tokens of embedded Topic/Subjects in copula clauses in Samir's contributions, and three more in the speech of Hocine and Najib. This is quite a large number if we consider that only 51 instances of Dutch NPs in MA clauses were found, even though Dutch insertions in MA are on average far more frequent. In the cases concerned here the clause is classified as a Dutch matrix because the copula and its verbal inflections are Dutch. The embedded MA Topic constituent is typically an independent pronoun, often a demonstrative.

I will first present embedded NPs that consist of a noun and a determiner. For another type of NP, a noun determined by a synthetic genitive, see *laylat l-qadr*, literally "the night of the divine decree", in example (2) above. I have eight examples of nouns prefixed by the MA definiteness marker *l-*: one by Najib, two by Hocine and five by Samir. In five of these, the MA constituent consists of the name of a language, which is marked by the definite prefix in MA. Dutch would not normally use a definite article in this particular context in informal speech, although it would sound overly formal rather than ungrammatical. Note the absence of an article before *franşawiya* in (20), which is in accordance with Dutch (ML) grammar.

- (17) *l-Şarabiya is heel mooi, snap je, heel aantrekkelijk*
 DEF-Arabic is very beautiful understand you very attractive
 "Arabic is very beautiful, very attractive." (Hocine)
- (18) *kijk, l-řanřa.. spanyuliya is een vreemd-e taal*
 look DEF-Fre.. Spanish is a foreign-AGR language
 "Look, Fre.. Spanish is a foreign language." (Hocine)
- (19) *l-ingliziya is niet zo internationaal als het eruit ziet*
 DEF-English is not as international as it looks-like
 "English is not as international as it seems." (Samir)
- (20) *l-ingliziya is eh veel eh staat veel hoger dan eh řanřawiya l-řan*
 DEF-English is er much er stands much higher than er French 'now
 "English is er much er has more status than er French now." (Samir)

Najib's (21) is different again: here the Dutch copula clause is coordinated with a verbal predicate in MA. (Alternatively, *is verstoep* can be regarded as an independent clause with an elliptic Subject.)

- (21) l-ŋqel dyal-hūm ma ka-y-fekker-š, *is verstoep*
 DEF-mind of-3PL NEG ASP-3-think-NEG is clogged
 "Their mind does not think, is clogged." (Najib)

In the next example *religie dyal-na* "our religion" is a MA constituent in my taxonomy because its internal structure must be attributed to MA. In this MA constituent, *religie* is an inserted Dutch noun. The NP *religie dyal-na* is the Subject of the Dutch copula *is* "is". Note also the MA sentence-initial Topic pronoun *hna*.

- (22) hna, *religie dyal-na is veel sterker eh in ons eh ingeprint, in ons hart*,
 1PL religion of-1PL is much stronger er in us er drummed in our heart
in onze hersens
 in our brains
 "As for us, our religion is much more er drummed into us, in our heart, in our brains." (Samir)

The remaining eleven instances all concern pronominal NPs. I quote some examples below. In most cases the MA constituent consists of a demonstrative pronoun, as in (23)-(25), and see also *had š-ši eh is onzin* "this is er nonsense" in (8) above.

- (23) la, hadak š-ši *is eh uit den boze*
 NEG DEM DEF-thing is er from the evil
 "No, this is fundamentally wrong." (Samir)
- (24) hadak hadak š-ši *is een verkeerde mentaliteit*
 DEM DEM DEF-thing is a wrong mentality
 "This, this is a wrong mentality." (Samir)
- (25) hadi *is één, één van de manieren*
 DEM is one one of the ways
 "This is one, one of the ways." (Samir)
- (26) wella, n-gul-u, ši wahed *is onredelijk wella dom*
 or 1-say-PL INDEF one is unreasonable or stupid
 "Or, suppose someone is unreasonable, or stupid." (Samir)

Twice Samir produces a 'hybrid' clause with both a MA and a Dutch copula. First in (27) the MA negation *maši* can be analysed as a negative copula "is not". *maši* is used only when there is no verbal or pronominal copula, that is, *maši* replaces the

‘zero’ copula in negative clauses.⁸ The negation and the copula are subsequently reiterated in Dutch. Then in (28) Samir first completes the clause as if there is a zero copula in MA, with the third person plural pronoun as the Subject/Topic on which the predication is made (*huma heel gevoelig* “they are very sensitive”). He subsequently rephrases in Dutch using the Dutch (3rd plural) copula verb *zijn*, but without repeating the Subject.

- (27) b *ṣ-ṣeħħ, hadi maši eh, is geen speciale cursus*
 with DEF-reality DEM NEG er is no special course
 “But this is not er not a special course.” (Samir)

- (28) *kayn-in ši ržal ta huma heel ge.. eh zijn heel gevoelig*
 EXIST-PL INDEF man·PL also 3PL very sens.. er are very sensitive
 “There are some men who are also very sens.. er are very sensitive.”
 (Samir)

9.1.2.3 Topic NPs in other than copula clauses

Most embedded MA pronouns occur as Topic and Subject in Dutch copula clauses, as we saw in the preceding section. Aside from these I have two examples. In (29) the indefinite pronoun *ši waħed* “someone” is the Subject of the Dutch matrix clause, as evidenced by its position followed by the Dutch finite verb, and by the agreement features of this verb. The first person pronoun *ana* in this example marks Topic shift, and is considered to be outside the ML clause (see the discussion in Chapter 3, sections 2.4.2 and 3.3.1).

- (29) *ana ma ʕemmɾ-i ši waħed heeft mij gediscrimineerd*
 1SG NEG ever-1SG some one has me discriminated
 “As for me, nobody ever discriminated against me.” (Hayat)

- (30) *hadak huwa vind ik interessant*
 DEM 3M find I interesting
 “This I find interesting.” (Samir)

Then *hadak huwa* in (30) has the syntactic function of DO of the verb *vind* “(I) find”. Being a topicalised (fronted) constituent in the Dutch clause, it triggers the inversion of the Subject and the finite verb. The 3M singular pronoun *huwa* is a discourse emphatic pronoun too, albeit of a different kind. (The emphatic pronoun that indicates a shift in Topic always occurs in the left-most position, while the emphatic pronoun that follows its referent (*hadak* in this case) merely expresses emphasis.)

⁸ Alternatively, it can be argued that *maši* merely negates the zero copula, and is not a copula itself.

9.1.2.4 Left-dislocated NPs, or clause-external Topics

In three instances, two by Samir and one by Abdellah, the MA Topic is realised as a left-dislocated constituent of the Dutch clause. The left-dislocated NP recurs in a Dutch resumptive pronoun *dat* (neuter singular, but often generalised to all inanimate referents) or *die* (plural and common gender singular, but often generalised to all animate referents). In the first two examples *dat* functions as the Subject of the Dutch copula clause. In Samir's (32) the MA constituent first recurs as the Subject pronoun *dat* and, subsequently in his rephrasing, as *daar* which is the form the non-human pronoun takes on when it is the complement of a preposition, in this case the clause-final *bij* "at".

- (31) hadak š-ši, *dat is wat anders*
 DEM DEF-thing that is something different
 "As for this, that's something else." (Abdellah)

- (32) u l-ingliziya *dat is .. daar is het niet bij*
 and DEF-English that is there is it not at
 "As for English, that's .. it doesn't pertain to it." (Samir)

Finally, consider (33) where a MA NP modified by a MA relative clause is realised as a left-dislocated constituent in a Dutch sentence. The Dutch pronoun *die* refers to the left-dislocated constituent.

- (33) had l-mğarba lli sakn-in hna-ya, *die moeten een betere*
 DEM DEF-Moroccan*PL REL live-PL here-EMPH these should a better

kans krijgen
 opportunity get
 "These Moroccans who live here, they should get a better opportunity."
 (Samir)

9.1.3 Summary

The investigation of nominal and adjectival MA insertions in Dutch yields some interesting observations and generalisations. Three main motivations emerge for embedding MA material in Dutch matrix constituents and clauses.

Firstly, there is the need for specialized terminology expressing Islamic and typically Moroccan concepts in order to fill lexical gaps in Dutch. This explains the majority of the embedded MA nouns. Sometimes the MA noun is prefixed with the definite article *l-*. In these cases, the MA article does not co-occur with a Dutch article, and the inflected MA noun can be classified as an embedded NP in view of its distribution in the Dutch clause. Where the MA definite prefix occurs, its selection is generally best explained by MA grammar, that is, the grammar of the EL. This is

not necessarily at odds with the MSA. The MSA states that EL constituents are well-formed according to the EL grammar, but the model makes no claims as to whether the ML of the clause selects the function morphemes that surface in the embedded constituent, or the language of the embedded constituent itself. Which types of EL constituents can occupy the position of a comparable ML constituent is ultimately a matter of ‘congruence’. The presence of MA articles in embedded MA constituents contrasts, however, with embedded Dutch NPs in MA clauses. In the latter case, Dutch function morphemes that have no clear correspondence to a MA counterpart are avoided, except for in the contributions of Hocine. We may conclude that the requirement of congruence poses few restrictions on the insertion of MA NPs in Dutch clauses, but more severely constrains the insertion of Dutch constituents that contain function morphemes, notably articles.

Secondly, it turns out that repetition as a means of enhancing discourse cohesion and marking topic continuity is a key factor in explaining the remaining embedded MA nouns and some adjectives. These embedded nouns and adjectives are not culturally specific like those in the first group just mentioned. Furthermore, it is striking that most examples occur in Dutch matrix constituents that are themselves embedded in a MA clause.

Thirdly, a large number of the embedded MA constituents consists of Topical NPs in clause-initial position. Most of these are independent demonstratives like *hadak š-ši* ‘this’. Also, in most cases they are Subjects, particularly of Dutch copula clauses. However, to identify them as Topics is a preferable generalisation because this includes example (30) where the Topic constituent is the DO of the ensuing clause, as well as the left-dislocated NPs in (33), (31) and (32) which are also not syntactically Subjects. These MA Topic NPs are clearly constituents in a Dutch syntactic matrix structure, so that the notion of insertion is also relevant here. The cases of left-dislocation trigger a resumptive pronoun in the following Dutch clause. The other Topic NPs occupy the first position of the Dutch clause, as evidenced by the occurrence of the Dutch finite verb in the second position (the so-called ‘verb second’ rule). In (30), where the Topic is not the Subject, the ‘verb second’ rule requires the inversion of the Subject and the finite verb. The fact that most embedded NPs are in Topic position in the Dutch clause may be the result of a tendency with some respondents to start an utterance in MA and subsequently to interrupt themselves and rephrase in Dutch, thereby incorporating the already pronounced element as first constituent in the Dutch clause structure.

Concerning the embedded NPs that are not in the first or left-dislocated position of the Dutch matrix clause, these are all MA nouns prefixed by the definite determiner. In these cases the presence of the determiner is generally better explained by MA grammar than by Dutch (ML) grammar.

It can be concluded that the concept of matrix language and insertion is still applicable when Dutch, the superimposed and culturally dominant language, is the ML. Moreover we see that the kinds of MA insertions are quite different from Dutch insertions in MA. Insertion of content words is relatively rare and considerations that

relate to discourse structure, notably cohesion through repetition and Topic-Comment sequences, are very prominent. These matters will be addressed in more detail in Chapter 11.

9.2 MA PPs and AdvPs in Dutch matrices

Following the organisation of Chapter 7, the embedded prepositional constituents will be dealt with in section 2.1, and the embedded MA adverbs in section 2.2 below. For general considerations regarding PPs and adverbs or adverbial constituents and their classification the relevant sections in Chapter 7 can be referred to. A summary concludes this section.

9.2.1 Embedded PPs

Nine embedded MA prepositional constituents occurred in the data. As in Chapter 7, I distinguish between complements and adjuncts. Three embedded PPs are locative adjuncts, three are analytic possessives with *dyal*. The examples that come closest to being complements are listed here first. A time adjunct, *men temmak*, literally “from there” will be discussed in section 2 on embedded adverbs (ex. (46) below). *b ş-şehh* in the sense of “but, however” will be discussed under adversative conjunctions in Chapter 11, although it is formally a PP, glossed literally as “with the reality”. No MA prepositional constituents were found to occur as modal adjuncts, or as predicates in Dutch copula clauses.

The clearest example of a complement PP was cited in section 1 above, and is repeated here as (34). The Dutch noun *twijfel* “doubt” subcategorizes for a PP complement with the preposition *aan*. In (34) Samir starts formulating the complement in Dutch, interrupts his flow and produces a PP complement in MA. The MA translation of “doubt”, *šekk*, plural *škuk*, actually subcategorizes for the selected preposition *f*.

- (34) *je had twijfels aan .. f d-din dyal-ek*
 you had doubts about in DEF-religion of-2SG
 “You had doubts about .. about your religion.” (Samir)

In (35) *b d-dariža* has the semantic role ‘instrumental’. The instrumental role is not an obligatory complement of Dutch *praten* “to talk”, but on the other hand instrumental arguments cannot occur as an adjunct with any verb and therefore share some of the characteristics of complement PPs. The selection of the preposition *b* to mark the language in which one talks is governed by MA. This is the appropriate preposition for this purpose in MA, whereas Dutch uses a locative *in*, as demonstrated by the coordinated Dutch PP *in het Arabisch* “in Arabic” in the same clause.

- (35) *ik zit met jou in het arabisch of eh b d-dariža te praten*
 I sit with you in the Arabic or er with DEF-Moroccan·Arabic to talk
 “I’m sitting here talking to you in Arabic or er in Moroccan Arabic.”
 (Abdellah)

The Dutch verb *vallen* “to fall” in (36) subcategorizes for a locative preposition to specify the time when something takes place. There is some freedom concerning the selection of the preposition and the preposition contributes itself to the meaning of the clause (cf. English “it falls before/after/on [a date]”). Therefore the PP *f sebša u šešrin* “on the twenty-seventh” is neither a prototypical complement nor a prototypical adjunct.⁹

- (36) *u durka hna-ya, wij gaan ervan uit dat ’t eh f sebša u šešrin valt*
 and now 1PL-EMPH we ‘assume’ that it er on seven and twenty falls
 “And now we assume that it [Laylat l-Qadr] falls on the twenty seventh [of Ramadan].” (Samir)

In examples (34)-(36) the selected preposition is the one we would expect were the subcategorizing verb or noun in MA.

The next three examples show locative adjuncts. The embedded PP *f l-meğrib* in (37) is a repetition, having already occurred in a MA clause in the two preceding speech turns by Samir and Fatima. (The preceding turn by Samir is reproduced as example (4) in Chapter 8, to which Fatima responds with *f l-meğrib*? “in Morocco?”.) In (38) the complement of the embedded preposition *f* “in” is the Dutch place name Tilburg.

- (37) *wah, wat bevalt je f l-meğrib?*
 yes what pleases you in DEF-Morocco
 “Yes, what do you like in Morocco?” (Samir)
- (38) *een boekbespreking, ik heb het gedaan f tilburg*
 a book·review I have it done in Tilburg
 “A book review, I did it in Tilburg.” (Hocine)

The third MA locative PP occurs in the Topic position of a Dutch clause, triggering the inversion of Subject and finite verbs in accordance with the ‘verb second’ rule.

- (39) *ana-ya, ana-ya, mša-k ben ik mezelf*
 1SG-EMPH 1SG-EMPH with-2SG am I myself
 ‘I, I, with you I’m being myself.’ (Samir)

⁹ See Chapter 7, note 3 on p. 274.

Finally, there are three instances of a MA possessive PP modifying a Dutch noun, two by Samir and another one by Hocine. In the examples below the preposition itself is the only MA morpheme; the third instance is *dyal remdan* in example (2) above.

- (40) *je hebt van die eh n-gul-u kern- eh kernproblemen dyal eh __ leven,*
 you have of these er 1-say-PL core er core-problems of er [the] life

ya-k

QTAG-2SG

“You have these er, let’s say, core, core problems of life, right?” (Samir)

- (41) *ik ben niet tevreden over eh __ kwaliteit dyal eh, ja, dyal __ faculteit*
 I am not satisfied about er [the] quality of er well of [the] faculty
 “I’m not satisfied with er the quality of er, well, of the faculty.” (Hocine)

Elsewhere I argued that in cases of layered embedding such as in (38)-(41), the most consistent analysis considers the entire PP as an embedded constituent and not just the preposition. From a crosslinguistic perspective on CS, the internal structure of the PP supports this analysis and, in particular, the relative order of adposition and its NP complement, this argument being apparent only when one language has prepositions and the other, postpositions (cf. Chapter 1, p. 47; Chapter 2, section 1.2.3). In the above examples (38) and (40), assuming that MA is the ML for the PP level helps somewhat to account for the absence of a definite article preceding the Dutch nouns *leven* “life” and *faculteit* “faculty”. In Chapter 5 we saw that, on the one hand, the MA definite prefix *l-* is often omitted before embedded Dutch while, on the other hand, embedded Dutch NPs seldom contain Dutch articles. Therefore the missing articles after *dyal* in (38) and (40) to some extent support the claim that we are dealing here with embedded PPs rather than embedded prepositions. However, the respondents of the Nijmegen corpus sometimes omit Dutch articles in monolingual stretches of Dutch as well, as the missing article before *kwaliteit* “quality” in Hocine’s (40) illustrates.

9.2.2. Embedded Adverbs and AdvPs

The MA adverbs in Dutch clauses are of two kinds: place and time adjuncts, and instances of *matalen* “for example” and *bhal daba* (ditto) which serve to qualify the utterances.

9.2.2.1 Place and time adverbs

There are nine instances of MA place and time adverbs in Dutch matrix clauses, all except one produced by Samir. Most of these occur in clause-initial (i.e., Topic) position, where they trigger the inversion of Subject and finite verb in the Dutch

clause. The MA word *hna* “here” and its emphatic variant *hnaya* are the only place adverbs. *Hna* occurs twice in clause-initial position:

- (42) *iyeh, hna gaan hun met de tijd mee*
 yes here keep·pace· they with the time ·keep·pace
 “Yes, here they keep pace with the times.” (Hayat)

- (43) *hna verveel je — kapot*
 here be·bored you [REFL PRON] broken
 “Here you’re bored to death.” (Samir)

The ‘emphatic’ form *hnaya* occurs twice as a modifier of a Dutch noun (cf. note 4 on p. 279 in Chapter 7). The use of the MA ‘emphatic’ suffix *-ya* after *hna* and personal pronouns (*ana-ya*, *nta-ya*, *nti-ya*, *hna-ya*) is so pervasive in Samir’s speech variety that I do not actually assume pragmatic emphasis here.

- (44) *die zijn verblind eh geraakt door eh hoe zeg je eh de rijkdom hna-ya*
 those are blinded er got by er how say you er the wealth here-EMPH
 “Those got blinded by er how do you say, er the wealth here.” (Samir)

- (45) *is het niet ook zo dat de mentaliteit misschien veranderd moet*
 is it not also so that the mentality perhaps changed must

worden bij de marokkanen hna-ya?
 become(PASS) among the Moroccans here-EMPH
 “Isn’t it also the case that perhaps the mentality has to be changed among the Moroccans here?” (Samir)

Three time adjuncts occur clause-initially, two clause-finally. The time adjunct *men temmak*, literally, “from there” in (46) is formally a PP, while *ši xetṛat* “sometimes” in (47) is formally an NP. (48) involves the rephrasing of a sentence started in MA, but without repeating the time adverb, which now occupies the Topic position in the Dutch clause (as shown by the inversion of Subject and finite verb).

- (46) *u men temmak zeṣma zeṣma beschouwde ik hem als een echte vriend*
 and from there EPIST EPIST considered I him as a real friend
 “And since then I considered him as a real friend.” (Samir)

- (47) *ši xetṛ-at is het beter om in beeldspraak te praten over iets*
 INDEF time-PL is it better COMP in metaphor to talk about something
dat niet duidelijk is dan..
 that not clear is than
 “Sometimes it is better to use a metaphor when talking about something that is not clear, than (..)” (Samir)

- (48) ka-t-šuf fisaʕ doorzie je alles
 ASP-2-see quickly see·through you everything
 “You see quickly, you see through everything.” (Samir)

Time adverbs typically occur directly after the finite verb in Dutch, unless they are topicalised, but the clause-final position is possible in cases where the adverb can be considered to be a kind of afterthought. One instance was cited in section 1: the Classical Arabic *l-ʔan* “now” in (20); (49), below, is the other one.

- (49) hoe lang doe je al stage durka?
 how long do you already apprenticeship now
 “For how long have you been doing this apprenticeship now?” (Samir)

9.2.2.2 The words *matalen* and *bḥal daba* “for example”

In the same way that Dutch *bijvoorbeeld* “for example” is used as a qualifying adverb in MA clauses in Fatima’s CS variety (cf. Chapter 7, section 2.6), instances of MA words for “for example” can be found both preceding and within Dutch clauses uttered by Jamal and, also, Fatima. Examples are presented hereafter. The pattern is very similar to that of the clause-initial discourse markers to be discussed later on in section 3.2: unlike the place and time adverbs of the preceding section, the clause-initial “for example”, whether Dutch *bijvoorbeeld* or MA *matalen*, does not trigger the inversion of Subject and finite verb in the Dutch main clause.

The word *matalen* “for example”, from Standard Arabic, is a common and widespread discourse marker in the somewhat more learned varieties of MA (it is pervasive in Najib’s MA utterances, for instance). It largely replaces the more dialectal *bḥal daba*, literally “like now”. Fatima also uses it in the context of Dutch clauses, although Dutch *bijvoorbeeld* occurs more frequently in her data:

- (50) matalen, ja, matalen, ja, ik heb daar vrienden, ja, da’s een kleine
 for·example yes for·example yes I have there friends yes that-is a small

voorbeeld
 example
 “For example, well, for example, well, I have friends there. That’s a small example.” (Fatima)

Jamal prefers the MA form *bḥal daba*. Consider the use of various Dutch and MA discourse markers in the following excerpt by Jamal, and note in particular *bḥal* “like, for instance” and *bḥal daba* in the first and third lines.

- (51) maar da’s wel interessant als je dit leest hè bḥal psychologie
 but that-is AFFIRM interesting when you this read QTAG like psychology

hè, is echt wel interessant hè, niet filosofie. ik ben het niet met
 QTAG is really AFFIRM interesting QTAG not philosophy I am it not with

jou eens hè, dak š-ši walu, walakin bħal daba psychologie
 you in*agreement QTAG DEM DEF-thing nothing but like now psychology

is echt interessant

is really interesting

“But it IS actually interesting if you read this. Like for instance psychology is really interesting; not philosophy. I don’t agree with you [that] this is all worthless. But for instance psychology is really interesting.” (Jamal)

9.2.3 Summary

The small number of tokens does not permit the making of strong claims; however, some tentative generalisations can be made. None of the embedded PPs gives an indication that Dutch, the ML on clause level, influences the selection of the preposition. The language of the governing head - at least the one that surfaces in speech - does not appear to influence selection of the complement preposition. When the examples discussed in this chapter are compared with the Dutch embedded adverbs and PPs in Chapter 7, it can be noted that no MA manner adverbs are inserted and that the modal uses that are prominent among the Dutch insertions are restricted here to MA words for “for example”. Another observation with respect to the MA adverbs is that they often occur in clause-initial, i.e. Topic, position in Dutch clauses, a tendency already established for MA nominal constituents in section 1 of this chapter.

9.3 MA Clauses in Dutch Matrices and Discourse Markers

The embedding of certain types of MA subordinate clauses in Dutch main clauses is discussed in section 3.1. After that, we will leave the realm of insertional CS. In section 3.2, I will demonstrate the use of various MA discourse markers in the context of Dutch clauses. While few Dutch conjunctions were found to conjoin MA sequences, MA conjunctions do link Dutch sequences. Section 3.3 is a summary.

9.3.1 Embedded clauses

The MA clauses that occur as embedded constituents are divided into four classes according to their function in the Dutch matrix clause: a) relative clauses (3.1.1), b) conditional clauses and temporal adjunct clauses (3.1.2), c) Subject clauses (3.1.3) and d) clauses that function as the complement of a verb, noun or adjective in a Dutch clause (3.1.4). In one instance, a MA clause functioned as the predicate of a Dutch

copula; this instance is presented in section 3.1.3, because the grammatical construction it involves is similar to that in Subject clauses.

I do not consider the MA clauses beginning with a causal conjunction as syntactically embedded. Samir, Maryam and Zineb occasionally switch between a Dutch clause and a MA clause that starts with a MA causal conjunction (*ŋla xaṭer* in Samir's or *ŋla heqqaš* in Maryam and Zineb's variant). These markers introduce an explanation in terms of a speech act justification ("I'm telling you this, because ...") or an inference ("I know this, because ..."), rather than a factual causation. (52) and (53) illustrate these uses. Note that in (52) *ŋla heqqaš* is part of the reported speech. Instead of providing the explanation announced by *ŋla heqqaš*, Zineb interrupts her flow and rephrases the reported speech in MA, providing the announced explanation in the second line, although this time without the causal conjunction.

- (52) *rūbbama y-gul-l-ek nta, jij bent heel anders, ŋla heqqaš eh*
 perhaps 3-say-to-2SG 2M you are very different because er

ka-y-gul-l-ek nta meḡribi f škel axūr, ka-t-t-ŋamel b waḥed
 ASP-3-say-to-2SG 2M Moroccan in shape other ASP-2-MP-treat with INDEF

ṭ-ṭariqa .. šwiya metherr-a
 DEF-manner little liberated-F

"Perhaps he'll tell you: 'YOU are very different. Because er ..' He tells you 'You are a different type of Moroccan, you behave in a .. somewhat liberated manner.'" (Zineb)

- (53) *ja, daar ben ik 't dus niet mee eens, la heqqaš fin saken huwa*
 yes that am I it so not with in-agreement because where live·PART 3M

dak l-meḡribi wella l-turki b l-wijk?
 DEM DEF-Moroccan or DEF-Turk in DEF-neighbourhood

"Yes, so I don't agree with this. Because: where in the neighbourhood does he live, this Moroccan or Turk?" (Maryam)

In the reported speech in (52) the (imagined) speaker uses *ŋla heqqaš* to clarify her or his point of view: "I think you are a different (i.e. acceptable) type of Moroccan, because you behave in a liberated (i.e. westernized) manner". The conjunction does not signal factual causation: "your liberated behaviour caused you to become different". Likewise in (53) *ŋla heqqaš* announces a clarification and argumentation, but not factual causation. In this example this is immediately clear as the conjunction is followed by a rhetoric question. In Chapter 3 I argued, following Schiffrin (1987) and Günthner (1993), that syntactic subordination and the grammatical characteristics associated with it apply only when factual causation is emphasized. The other types of causal relations, i.e. Warrant-Inference and Motive-Action in Schiffrin's terminology, are best explained in terms of discourse grammar. See the discussion

of example (1) in Chapter 3, section 1 and section 2.3.2 of that chapter; the issue will be taken up once again in section 3.2.1 below on causal conjunctions.

9.3.1.1 Relative clauses

For a characterisation and comparison of relative clauses in MA and Dutch see section 1.1 in Chapter 8. An interesting question with regard to embedded relative clauses is whether they are marked by a relative pronoun as in Dutch, or by the particle *lli* as in MA. There are only three examples of MA relative clauses embedded in Dutch matrix clauses, all of them different. The relative clause in (54) is perfectly well-formed according to MA grammar: it contains the relative clause marker *lli* which is obligatory when the head is definite. In (55), however, the MA relative clause begins with the Dutch relative pronoun *die*. The MA particle *lli* is absent, although its use is optional with an indefinite head NP. (This example shows an agreement failure: while the head noun *dingen* “things” and the relative pronoun *die* are plural, the MA verb form *ka-ye ſzeb* “it pleases” is singular.) The relative clause in Younes’ (56) is not formally marked as such, but it is well-formed according to MA grammar, because the head NP *heel veel* “very many” is indefinite.

- (54) *je hebt niet je eh hoe zeg je dat, de wens lli kan-et ſend-ek bekri*
 you have not your er how say you that the wish REL be-3F at-2SG before
 “You don’t have your er, how do you say this, the wish you used to have?”
 (Samir)
- (55) *nog meer dingen van ř-řyafa die ma ka-ye-ſzeb-l-ek-ši?*
 still more things of DEF-Rifian·PL that(REL PRON) NEG ASP-3-please-to-2SG-NEG
 “[Are there] any more things about Rifians that you don’t like?” (Samir)
- (56) *ik ken er heel veel zeſma ma y-dir-u-š aanpass-en*
 I know EXPLETIVE very many EPIST NEG 3-do-PL-NEG adapt-INF
 “I know very many who don’t adapt [themselves to Dutch culture].”
 (Younes)

9.3.1.2 Conditional clauses and temporal adjuncts

While there were five Dutch conditional clauses embedded in MA matrices (section 1.2 of Chapter 8), there are 11 examples of the opposite case: MA conditions in Dutch clauses. The distribution among the respondents is as follows: Samir 4 tokens, Abdellah 3, and two tokens each for Fatima and Hocine. The MA conditional clauses are usually marked by the conjunction *ila*, which is followed by the perfect verb in some varieties of MA, and by the imperfect in others. The conditional clause may also lack any formal marking. The MA conditional clause occupies the first constituent position in the Dutch main clause, thus triggering the inversion of the finite

verb and the Subject in the main clause. Alternatively the Dutch main clause may be introduced by the Dutch particle *dan* “then”. Consider the examples listed hereafter.

- (57) u ila dewwez-ti l-ṣuṭla te-rṣeṣ hna, *dan heb je hetzelfde probleem*
 and if pass-2SG DEF-holiday 2-return here then have you the·same problem
 “And if you pass your holiday [in Morocco] and you come back here, you have the same problem.” (Abdellah)

- (58) b ṣ-ṣeḥḥ ila te-qra ši ḥaḥa lli ma ka-te-ṣṣeb-l-ek-š,
 with DEF-reality if 2-study INDEF thing REL NEG ASP-3F-please-to-2SG-NEG

waxxa ġadi t-zid te-nṣeḥ fi-ha, *ben je niet gelukkig*
 although FUT 2-proceed 2-be·successful in-3F are you not happy
 “But if you study something you don’t like, even if you’re successful in it, you won’t be happy.” (Samir)

- (59) ila der-t-l-ek ana l-ṣerḍ u t-hezz-u nta te-ṣṣi-h-l-u, *is dat*
 if do-1SG-for-2SG 1SG DEF-paper and 2-carry-3M 2M 2-give-3M-to-3M is this

duidelijk

obvious

“If I write the paper for you and [then] you give it to him, this [i.e. the fraud] will be obvious.” (Hocine)

- (60) ila ka.. bḡi-ti te-qra ṣla ši ḥaḥa, *ja, muhimm, bijvoorbeeld*
 if ASP- want-2SG 2-study about INDEF thing yes anyway for·example

bḡi-ti t-dir-i ši *jouw eigen zaak ofzo, hè, moet je*
 want-2SG 2-do-F INDEF your own business or·something QTAG must you

zoveel voor leren hè (u je moet __ zoveel over leren)

so·much for learn QTAG and you must [it] so·much about learn

“If you want to learn about something, well, anyway for example, you want to have your own business or something, right? you have to study so much for this, you have to learn so much about [it].” (Fatima)

In addition to the conditional clauses, I note two instances of embedded MA clauses that function as temporal adjuncts. Like the conditions, they occupy the first position in the Dutch main clause and lead to the inversion of Subject and finite verb in Dutch.

- (61) melli bda naḥib *was hij de enigste*
 when start Najib was he the only·one
 “When Najib started, he was the only one.” (Hocine)

- (62) la būdd ma ṣad t-kun l-wedṣiya methessn-a, *natuurlijk hè*, *dan*
absolutely when 3F-be DEF-situation improved-F of-course QTAG then

ga ik met hem mee terug
accompany I with him accompany back

“Absolutely, when the situation has improved, of course, then I will go back with him.” (Fatima)

9.3.1.3 Subject clauses

While no Dutch Subject clauses are embedded in MA matrix clauses, MA clauses do occur as Subjects in Dutch matrices. Examples are found with Samir (2 tokens) and Fatima (2 tokens). In Dutch, Subject clauses typically occur at the end of the main clause, and they are represented by the expletive pronoun *het* ~ *'t* “it” in Subject position. The Subject clause itself is introduced by the complementizer *dat* “that”. This is the syntactic construction found in (63). Note that in this example the Subject of the MA Subject clause is expressed in Dutch as the pronoun *-ie*, an allomorph of *hij* “he” that is cliticized to the complementizer.

- (63) 't kan niet dat-ie eh b z-zezz ye-ržeṣ meslem
it can not that-he er with DEF-force 3-become Muslim
“It’s not possible that he’s forced to become a Muslim.” (Samir)

Fatima’s (64) contains neither a Dutch expletive pronoun nor a complementizer. It can be argued that the MA presentative particle *ra-* and the suffixed pronoun serve the function of expletive. The modal meaning of the main clause (“it’s not necessary”) is reflected in the copula in the MA Subject clause. In order to express this modality the subjunctive form of the verb *kan* “to be” replaces the zero or pronoun copula.

- (64) ra-h ma .. *hoeft niet persé* y-kun eh t-kun mṣaḥb-a mṣa-h
PRES-3M NEG is-necessary not absolutely 3-be er 2-be befriended-F with-3M
“It’s not, it’s not absolutely necessary that he’s er .. you’re befriended with him.” (Fatima)

A similar and even more intriguing case is found in (65) where the scope of the Dutch modal auxiliary *hoeft niet* “doesn’t have to”¹⁰ extends to include the Dutch infinitive as well as the conjoined MA clause. In the MA clause this is again signalled morphologically by the subjunctive mood (i.e., without *ka-*) in *yšekk* “he doubts”. From the viewpoint of MA grammar the MA clause *yšekk f ši ḥaža* “he has doubts about something” could be analysed as the Subject of the verb “not to be necessary”;

¹⁰ This modal verb *hoeven* occurs only with a negation or negative-like element in Dutch; it is the common negation of *moeten* “to need, have to”.

from the viewpoint of Dutch grammar, the pronoun *hij* ‘he’, referring to a male person, is the Subject of *hoeft niet*.

- (65) *dus hij hoeft niet bang te zijn of y-šekk f ši ɣaža*
 so he needs not afraid to be or 3-doubt at INDEF thing
 “So he doesn’t have to be afraid or suspicious about anything.” (Fatima)

In the fourth example, the MA Subject clause occurs in the place of a Dutch clause of the type [*om + te + infinitive*]. The MA equivalent of this type of clause is marked by the complementizer *baš* followed by the subjunctive verb. Similar examples are discussed under complement clauses below. The Dutch clause of which *baš tetbe f ši ɣaža hnaya* in (66) is the Subject is itself the complement of *ik weet* ‘I know’.

- (66) *ik weet dat het moeilijk is baš te-tbe f ši ɣaža hna-ya*
 I know that it difficult is COMP 2-follow INDEF thing here-EMPH
 “I know that it is difficult to follow something here.” (Samir)

Finally, consider example (67). Here the MA clause is syntactically the predicate of the copula *is*, which has *’t verschil* ‘the difference’ as its Subject. Although the syntactic function of the MA clause is formally different, the syntactic construction is quite similar: the MA clause is introduced by the Dutch complementizer *dat*, but it has no dummy pronoun in Subject position.

- (67) *’t verschil is wel dat eh l-huḷandīya hiya l-luḡa lli*
 the difference is AFFIRM that er DEF-Dutch 3F DEF-language REL

t-ṣellem-t-ha n-gul-u rebṣataš sana
 MP-teach-1SG-3F 1-say-PL fourteen year
 “The difference is that er Dutch is the language that I learnt for fourteen years.” (Samir)

9.3.1.4 Complement clauses

Setting aside reported speech as the complement of a verb of saying, there are six examples of embedded MA clauses that function as a complement in a Dutch main clause. Three of these are interrogative clauses; in (69) and (70), the question is fronted to the first position (note the inversion of Subject and finite verb in Dutch).

- (68) *zij moet weten šhal men ṣam ṣend-ki u ..*
 she must know how many of year at-2F and
 “She has to know how old you are and ..” (Samir)
- (69) *mnin ži-ti hoef je eigenlijk niet te zeggen*
 from where come-3SG need you really not to say
 “Where you come from, you don’t really need to tell.” (Samir)

- (70) *Ťlaš ma ži-ti-š, eigenlijk, vraag ik me af, ya-k?*
 why NEG come-2SG-NEG actually wonder I me-REFL wonder QTAG-2SG
 “Why didn’t you come, actually, I wonder.” (Jamal)

The complement of *ik denk* “I think” in (71) below is a declarative clause. In Dutch, embedded (in the syntactic sense of the word) declarative clauses are marked by the complementizer *dat*, which is not realised in this example. However, an unmarked complement like *huwa b l-huḷandīya* “this is in Dutch” is grammatical in MA, although a complementizer like *belli* would also be possible here.

- (71) *ik denk huwa b l-huḷandīya, ik denk ’t wel*
 I think 3M with DEF-Dutch I think it AFFIRM
 “I think this is in Dutch, I really think so.” (Abdellah)

The remaining two instances are MA clauses of the type [*baš* + subjunctive verb] which again occur in the place of the Dutch type [*om* + *te* + infinitive]. The congruence of these two clause types has been shown in example (66) above and in Chapters 5 and 6, where MA ‘*baš*-clauses’ were discussed as complements of embedded Dutch nouns and verbs.

- (72) *waš zeŤma huma eh zijn zij in staat baš y-qerri-w l-Ťarab ?*
 Q EPIST 3PL er are they in state COMP 3-teach-PL DEF-Arab’s
 “Are they, well, are they capable of teaching Arabs?” (Samir)

- (73) *ik had eh geen zin baš n-dir telt snin wella twee jaar alleen*
 I had er no liking COMP 1-do three year-PL or two year just

— *de propaedeuse*

[for] the propaedeutic-diploma

“I didn’t feel like doing three years or two years just [for] the propaedeutic diploma.” (Hocine)

Reported speech as a special type of embedded clause occurs with both Dutch and MA as the matrix language (see section 1.4 in Chapter 8). As I pointed out in Chapter 8, any string of words can in principle occur as the complement of a verb like “to say”, “to think” et cetera. Since the constraints are barely syntactic in nature, the phenomenon of reported speech is not investigated in this corpus description. Again, I will just give a few illustrations.

- (74) *stel je voor: ġadi te-lqa ši xedma mezyan-a f l-meġrib*
 imagine you-REFL imagine FUT 2-find INDEF job good-F in DEF-Morocco
 “Imagine: you’ll find a good job in Morocco.” (Samir)

- (75) *je hoort ook in marokko dat ze zeggen:* ka-ye-t-fehbm-u¹¹
 you hear also in Morocco that they say ASP-3-MP-make-understand-PL

ʕli-na

over-1PL

“You also hear in Morocco that they say: ‘They think they’re better than us.’
 ” (Maryam)

9.3.2 MA conjunctions and other discourse markers

The following discussion of MA discourse markers that are used in the context of Dutch clauses will not be exhaustive because, as was argued in Chapter 3, these phenomena fall largely outside the framework of insertional CS. Only the patterns of use that are most noticeable in the data corpus will be presented here. One type of discourse marker, viz. *matalen* “for example” and *bḥal daba* (ditto), was addressed in section 2.2.2 above. The use of discourse emphatic pronouns has been discussed at length in sections 2.4.2 and 3.3.1 of Chapter 3 and will not be reiterated here. Some further examples of MA clause-initial Topic pronouns are found in the examples cited in this chapter: *nta* in the first line of (52), *huma* in (72) above and *anaya* in the second line of (86) below. See also (29), (22) and (36) above.

Hereafter I will demonstrate the use of three types of MA conjunctions that recurrently introduce Dutch clauses, conjoining these to a preceding clause or another sequence of discourse in either Dutch or MA. These three types of conjunctions can be characterised as causal “because”, adversative “but” and coordinate “or” and “and”. The status of these conjunctions as embedded words in a Dutch matrix clause or sentence is not so obvious. It can be argued just as convincingly that the conjunction and the ensuing clause are both constituents in a larger matrix structure (‘sentence’, or CP in approaches inspired by X-bar theory) that is attributable to MA grammar. Refer to Chapter 3 where this discussion is worked out in detail. Following the discussion on conjunctions, I will treat the use of the MA question marker *waš* that incidentally precedes yes/no questions in Dutch.

Other MA discourse markers that are used in Dutch contexts include the ‘sequence-initial’ word *muhimm* “anyway” (from Standard Arabic) that is used in attempts to conclude the discussion of a topic (cf. example number (60) above, where it precedes a MA clause), *ngulu* “let’s say”, a common filler word in Samir’s utterances, and *ze ʕma*, which has appreciative modal values in MA (Caubet n.d. [1995]) but may be used as a filler as well in Samir’s variety. *ngulu* and *ze ʕma* occur in hesitation contexts. Clause-finally we find the question tag *ya-k*, an instance of which is cited

¹¹ The expected form would rather be *ka-yfehmu ʕli-na*, literally “they understand better than us”. According to Maryam, the Moroccans in Morocco complain about the arrogance of their compatriots in the European diaspora.

below. Interestingly, though, it is much more common to find the Dutch question tag *hè* attached to MA clauses (see Chapter 8, section 2.2) than MA *ya-k* to Dutch clauses. This is also true for Jamal's contributions, from which (76) is taken. See (40) above for another example.

- (76) *dat zijn altijd zo smoesjes, ya-k?*
 that are always such excuses QTAG-2SG
 "Those are always kind of excuses, aren't they?" (Jamal)

9.3.2.1 Causal conjunctions

Samir uses MA *ʕla xaṭeṛ ~ la xaṭeṛ* "because" to introduce explanations in Dutch. This appears rather frequently in Samir's data (more than 10 occurrences). In addition, a single instance of *li ʔanna* "because" was produced by Hocine. Examples are given below.

- (77) *ya laṭif! dat is een interessant probleem ʕla xaṭeṛ eh je leert een*
 Goodness! that is a interesting problem because er you learn a
tweede taal d'r bij, terwijl het niet je eigen moedertaal is
 second language it next-to while it not your own mother-tongue is
 "Goodness! That's an interesting problem, because er you learn a second
 language next to it, while it's not your own mother tongue!" (Samir)
- (78) *had š-ši eh hypocrisie, ʕla xaṭeṛ ik ben helemaal niet zo*
 DEM DEF-thing er hypocrisy because I am at'all not like-this
 "This is er hypocrisy, because I'm not like this at all." (Samir)
- (79) *t-kun-u mʕa ž-žuž hna-ya, la xaṭeṛ ik moet jullie handtekeningen*
 2-be-PL with DEF-two here-EMPH because I must your(PL) signatures
persé hebben
 absolutely have
 "You be here around two, because I really need to have your signatures."
 (Samir)

In the first two examples, *ʕla xaṭeṛ* introduces an explanation in terms of inference ("I know / say this, because .."); in (79) we have another clear example of "because" to mark the justification of a speech act ("I'm telling you this, because .."), as in the example quoted as (1) in Chapter 3. I have no example where *ʕla xaṭeṛ* is used to introduce a Dutch clause that unequivocally expresses factual causation. For factual causation, Dutch typically uses the conjunction *omdat* plus subordinate clause (i.e. verb-final) word order. The following example by Hocine comes closest to factual causation, but the Dutch clause has main clause word order. Hocine uses the

conjunction *liʔanna* “because”, which is a loanword from Standard Arabic but common in educated variants of MA.

- (80) ʔref-ti ʔlaʃ? liʔanna *het komt uit zijn onderzoek*
 know-2SG why because it comes from his research
 “Do you know why [he knows all this]? Because it comes from his research.”
 (Hocine)

9.3.2.2 Adversative conjunctions

Of all MA conjunctions conjoining Dutch clauses, the adversative ones, meaning “but, however”, are the most frequent and widespread in the Nijmegen corpus. They come in three (dialectal) variants: *walakin*, *walayenni*, and, in Samir’s eastern variety of MA, *b ʃ-ʃeħħ*. Examples occur in the contributions of Fatima, Samir, Jamal, Hocine, Hayat and Maryam; however, the phenomenon appears most frequently in the speech of the former two.

- (81) *ja ik wil graag terug walakin dat kan niet*
 yes I want eagerly back but that can not
 “Yes, I do want to go back, but that’s impossible.” (Fatima)
- (82) *kayen, kayen šwiya men dak š-ši walayenni dat is niet de eh echte*
 EXIST EXIST bit of DEM DEF-thing but that is not the er real
reden
 reason
 “Sure, there is some of this in it, but that’s not the real reason.” (Abdellah)
- (83) *dat is eh heel in harmonie met je ouders plus eh — buitenwereld,*
 that is er very in harmony with your parents plus er [the] outside-world

walakin dit hè, dat is niks, wat wij doen
 but this QTAG that is nothing what we do
 “That is totally in harmony with your parents plus the outside world, but this, you know, that’s nothing, what we do.” (Samir)
- (84) *(kayn-in ka-y-šerɓ-u š-šrab wella y-dir-u l-ħešš u eh)*
 EXIST-PL ASP-3-drink-PL DEF-alcohol or 3-do-PL DEF-hashish and er

ik wil dat niet goedpraten, b ʃ-ʃeħħ het heeft een reden, waarom
 I want that not justify with DEF-reality it has a reason why

je begint
 you start

“There are some who drink alcohol and use hashish and er I don’t want to justify it, but there is a reason, why you start.” (Samir)

- (85) *ik heb mijn eigen bankrekeningnummer, walakin mijn man, als hij*
 I have my own bank-account-number but my husband when he

geld wil opnemen ..

money wants withdraw

“I do have my own bank account, but my husband, when he wants to withdraw money ..” (Maryam)

- (86) M *is hij meer geïntegreerd dan jij?*
 is he more integrated than you
 “Is he more integrated [in Dutch society] than you?”

H *iyeh, walakin eh ana-ya, ik pas overal*
 yes but er 1SG-EMPH I fit everywhere
 “Yes, but er I fit anywhere.”

(Hayat and her daughter Maryam)

In the latter example by Hayat, MA *walakin* “but” precedes the MA emphatic pronoun *anaya*. The pronoun indicates Topic shift from HE (i.e., Hayat’s husband) to I (Hayat). This makes it difficult to argue that *walakin* is embedded in a Dutch matrix, since *anaya* would have no place in a Dutch matrix (cf. Chapter 3).

9.3.2.3 Coordinate conjunctions

There are some instances of the MA coordinate conjunctions *u* “and” and *wella* “or” introducing Dutch clauses, although this seems to be less frequent than adversative *walakin* and *be-ṣ-ṣeḥḥ*. Some examples follow again (see also *u* in the last line of Fatima’s (60) and the first line of Samir’s (84) above). Below, consider *wella* “or” in the second line of (87), and *u* “and” in (88) and (89).

- (87) *u ṣad te-bqa waḥed probleem dyal ‘je moet zoveel diploma-s*
 and still 3F-remain INDEF problem of you must so-many diploma-PL

hebben om dat te doen’ ja, wella je hebt geen moeite met diploma’s
 have to that to do yes or you have no difficulty with diplomas

enzo, dan heb je problemen met geld

et cetera then have you problems with money

“And there’ll always remain a problem of ‘you need so many diplomas to do that’. Yes. Or: if you have no difficulty with diplomas et cetera, then you have problems with money.” (Fatima)

- (88) *šhal t-šeddeb baš y-žib-na u ik ben hier nu vijf jaren,*
 how-much MP-make-suffer COMP 3-bring-1PL and I am here now five years
u šad n-welli n-gul-l-u eh ..
 and finally 1-become n-say-to-3M er
 “He suffered so much to bring us here, and I’ve been here for five years now,
 and finally I tell him er [I want to go back].” (Fatima)
- (89) *kayen hadak l-ferq u die kun je niet eh*
 EXIST DEM DEF-difference and this can you not er
 “There is this difference, and this you can’t er .. [interrupted].” (Samir)

9.3.2.4 The interrogative particle *waš*

An interesting pattern results from the use of the MA particle *waš* preceding questions in Dutch. This particle marks yes/no questions in MA; in Dutch such questions are normally marked by the inversion of the finite verb and the Subject, so that the verb appears in clause-initial position. In both languages questions are signalled by a rising intonation. When *waš* precedes a Dutch question, Dutch verb-initial word order applies as well, so that the question becomes doubly marked. This CS pattern, illustrated below, is found with Samir, Hocine and Fatima.

- (90) *wat bedoelt-u eh waš eh mag zij niet van haar ouders naar school,*
 what mean-you er Q er is-allowed she not by her parents to school
of van haarzelf?
 or by herself
 “What do you mean, er, is it her parents who don’t allow her to go to school,
 or herself?” (Fatima)
- (91) *n-gul-u-l-u waš is dit een goed verhaal?*
 1-say-PL-to-3M Q is this a good story
 “We’ll ask him: ‘Is this a good story?’ ” (Hocine)
- (92) *waš zijn ze hard wella wašta, wašta ka-t-gul nta?*
 Q are they tough or what what ASP-2-say 2M
 “Are they tough or what, what do YOU say?” (Samir)
- (93) *u nti-ya eh zešma kan-et eh waš is het een eh gedwongen keuze?*
 and 2F-EMPH er EPIST be-3F er Q is it a er forced choice
 “And as for you, it was so to say er, is it a forced choice?” (Samir)

The pattern illustrated here probably results from a rephrasing of the question in Dutch. This is apparent in the next example:

- (94) *waš hadik er n-gul-u is dat aan hunzelf te wijten*
 Q DEM er 1-say-PL is that on themselves to blame
 “Is this er, let’s say, do they have themselves to blame for that?” (Samir)

9.3.3 Summary

The insertion paradigm proved to be a valid approach to the study of subordination in CS. The embedding of MA clauses in Dutch main clauses has syntactic consequences that are adequately described in terms of ML structure and EL constituent. This is well illustrated by the MA conditional clauses that function as the first constituent in the Dutch clause, and hence trigger the inversion of the Subject and the inflected verb in the main clause. Likewise, MA complement clauses clearly assume a syntactic position in the Dutch ML clause.

Certain types of MA discourse marking are conspicuous in the context of Dutch clauses. In addition to the well-known phenomenon of clause-initial Topic pronouns that was discussed at length in Chapter 3, a number of conjunctions have been noted. These causal, adversative and coordinate conjunctions structure discourse sequences rather than conjoin clauses. Generally, as was argued in Chapter 3, discourse markers in CS are not very successfully dealt with in an insertion framework, and no such attempt was made in this chapter. The MA question marker *waš* is a good example of a discourse marker that is hard to conceive of as an inserted element since, after all, the Dutch matrix frame does not have a slot for question marking particles.

The MA discourse markers discussed in section 3.2 share the property of being ‘first’ elements with many of the embedded MA constituents discussed earlier. The clause-initial nominal and adverbial constituents are syntactically incorporated as Topic constituents in the Dutch matrix clause; these discourse markers are the first element of a ‘sequence’ or ‘unit of talk’ as defined by Schiffrin (cf. Chapter 3, section 1). The MA discourse markers are not syntactically embedded as first constituent in the Dutch clause (neither are many Dutch discourse markers, for that matter).

9.4 Individual variation among respondents

In connection with their personal migration history, the respondents vary with respect to the language they use most in everyday life and the language they speak most fluently (Chapter 4). This also influences their CS patterns. The insertion of Dutch content words in MA constituents is found with all respondents, however it is particularly frequent in the speech of the second generation Moroccans in the Netherlands. Dutch insertions in the contributions of Warda and Mustafa are very few; in the case of Mustafa, whose competence in Dutch was very limited at the time, embedded Dutch words are clearly linked to the two social settings where he used

Dutch: his Dutch language course, and another course where he learnt to repair typewriters and sewing machines and to weld boilers. The respondents who use more Dutch in their daily lives also use more Dutch in longer stretches during the conversations. This leads, among other things, to inter-sentential switching within as well as in between speech turns, aspects of CS which fall outside the scope of the present study. The same respondents who use longer stretches of Dutch during the recordings not only insert more Dutch content words, but also insert more complex Dutch constituents such as NPs, PPs and subordinate clauses. For more details on intergenerational differences in CS patterns in the Moroccan and Turkish communities in the Netherlands, see Backus & Boumans (1996) and Backus (1996a; 1996b: 134-6, 387-91). Not all idiosyncrasies are readily associated with sociolinguistic variables, however.

Much variation was found with respect to the insertion of Dutch verbs. The MA periphrastic *do*-construction was observed to be frequent in the speech of the Hamadi brothers and Hocine, infrequent in the speech of another five respondents, and absent from the CS varieties of five others. It should be noted that there is no direct relationship between the use of *dar* plus Dutch infinitive and competence in Dutch or the overall frequency of CS in the idiosyncratic CS varieties. The periphrastic construction is found in the contributions of respondents who had poor fluency in Dutch (Mustafa), as well as Dutch-dominant respondents (the Hamadi siblings); conversely, the construction is absent from the CS behaviour of some Dutch learners (Warda) as well as some fluent and Dutch-dominant bilinguals (Abdellah, Maryam).

Conventionalisation of CS patterns is most observable in the contributions of the Hamadi siblings, as the idiosyncratic varieties of all four siblings have much in common. This is particularly apparent in the use of the *dar* plus infinitive construction by the three Hamadi brothers. (Their sister Nawal's variety is not demonstrably different; she simply produced too few instances of this construction to allow for a comparison.) The Hamadi siblings may be compared to Maryam, who like them grew up in the Netherlands and has better mastery over Dutch than Moroccan Arabic. Maryam speaks Dutch most of the time during the recording sessions, even though her interlocutors speak mainly MA. Accordingly, she does not codeswitch much nor does she use the *dar* plus infinitive construction in order to embed Dutch verbs. This may be related to the fact that Maryam is not a member of a speech community where MA/Dutch CS is a common mode of communication, or at least not as common as with the Hamadi siblings. The latter report to engage in MA/Dutch CS with peers of Moroccan descent; conversely, Maryam and her mother Hayat report that they were the only Moroccan family in their village and that they had little contact with compatriots.

Remarkably, idiosyncratic patterns were found in the contributions by Fatima and Hocine. Fatima is the only respondent to use many Dutch modal adverbs in otherwise MA contexts. Hocine's bilingual speech diverges from that of other respondents with a comparable immigration history like Zineb, Najib, Mimoun and Fatima. Hocine's variety displays more Dutch insertions both quantitatively and qualitatively, in terms

of insertion types, despite a relatively low competence in Dutch (cf. *Table 4.2* in Chapter 4). Most conspicuously, he inserts Dutch NPs and PPs far more frequently than any other respondent, including those born and raised in the Netherlands. In addition, he is the only one to apply the MA word order Verb-Object when inserting Dutch infinitive verbs together with a Dutch lexical complement. We might say that Hocine is a more ‘daring’ individual when it comes to codeswitching. His divergent linguistic behaviour may also be related to his specific language abilities: being a native speaker of Tarifit Berber, he has less than native speaker fluency in both Moroccan Arabic and Dutch. He shares this feature with the other Tarifit speaker Najib, but Hocine’s divergent CS patterns have no obvious parallels in Najib’s data, as the latter does not codeswitch much at all. (Concerning the Tashelhit speakers Abdellah, Mimoun and Jamal, their MA is not readily distinguishable from that of the Arabophone respondents.) The matter of fluency or competence in MA requires qualification: when we compare the MA of Najib and Hocine to the variety spoken by the younger Hamadi siblings, we will see that the former have a much larger learned vocabulary at their disposal. Najib and Hocine are both skilled in Standard Arabic and draw from written Arabic sources to fill up their learned vocabulary in Moroccan Arabic and, probably, in Tarifit as well. For the Dutch born Hamadi siblings, Dutch is the primary source for learned vocabulary. However, upon close investigation, the Hamadi children may turn out to have more knowledge of idiomatic vocabulary pertaining to everyday life and, more importantly, certain grammatical structures of MA may prove to be more stable and more automatically produced (cf. Chapter 4, n. 2 on p. 163 on the Tarifit speakers’ competence in MA).

9.5 Overview table

In order to provide a quick overview of the insertion types that occur with either matrix language, the main facts of Chapters 5 to 9 are summarised in a table. A final summary of the findings forms part of Chapter 10.

	Dutch insertions in MA	MA insertions in Dutch
Nouns	hundreds of instances; variable gender assignment; Dutch plural markers; MA definite prefix <i>l-</i> often omitted before Dutch noun, as well as preposition <i>l</i> ; occur in all syntactic positions but not modified by MA adjective.	7 tokens referring to highly specific Islamic or Moroccan concepts; 4 tokens in repetition contexts; assigned common gender; no instance of plural forms.

	Dutch insertions in MA	MA insertions in Dutch
Attributive Adjectives	very rare; 2 unequivocal examples follow MA head noun (= MA, ML order).	1 instance in repetition context; follows Dutch noun (MA, EL order).
Predicative Adjectives	> 100 tokens; often modified by Dutch adverb, never by MA adverb; sometimes with Dutch or MA PP complement; examples of comparative and superlative uses.	1 token in repetition context; modified by Dutch adverb.
NPs	51 tokens distributed over various NP types: determined by Dutch article, possessive pronoun, numeral or other quantifier; relatively frequent a) as predicate of copula constructions, b) in the speech of Hocine, c) with numeral or other quantifier.	30 tokens; 5 of these culturally specific nouns accompanied by definite article; others clause-initial Topics, mostly in copula clauses; 3 of these left-dislocated.
Verbs	205 tokens in periphrastic <i>do</i> -construction <i>dar</i> + infinitive; much quantitative and qualitative variation among respondents: frequent with some, absent with other speakers; pronominal complements realised as suffixes on <i>dar</i> ; lexical complements always in Dutch, and following Dutch word order, except for Hocine; clausal complements in either language; Dutch reflexive pronouns with inherently reflexive verbs.	not attested
PPs	44 tokens (excluding discourse markers and complements of embedded Dutch words, notably verbs); 7 complements, 24 non-modal adjuncts, 9 modal adjuncts, 4 predicates of copulas.	9 tokens: 3 complements, 3 possessive, 3 locative, 1 of which occurs in Topic position.

	Dutch insertions in MA	MA insertions in Dutch
Adverbs	38 tokens (excl. modifiers of embedded words, adverbially used Ns and clock times, and <i>bijvoorbeeld</i> “for example”) these include 8 frequency and degree adverbs, 18 manner, 12 modal; can mostly be classified as AdvPs, do not modify MA adverbs or adjectives; frequent with Fatima (modals) and Hocine (other types).	9 time and place adjuncts, of which 5 occur in Topic position; in addition <i>matalen</i> , <i>bḥal</i> and <i>bḥal daba</i> “for instance” as discourse markers (not counted).
Subordinate Clauses	22 tokens (excl. complements of embedded verbs, and reported speech): 1 nominalised clause, 5 relative, 5 conditional, 6 complements, 5 adverbial adjuncts; mostly in Samir’s contributions.	27 tokens (excl. reported speech): 3 relative, 11 conditional, 2 temporal adjuncts, 6 complements, 4 Subject, 1 predicate clause. Many occur as 1st constituent of the Dutch matrix clause.
Discourse markers (impressions)	<i>ja</i> “yes, well” and <i>nee</i> “no”; question tag <i>hè</i> ; some adverbs and PPs; rare instances of conjunctions.	adverbs meaning “for instance”; apart from these, predilection for clause-initial position: discourse-emphatic pronouns, causal and adversative conjunctions, some coordinate conjunctions, question marker <i>waš</i>

Table 9.1. Comparison of insertion types in matrices from either language

Part III

Conclusions and Outlooks

wa ʕinda al-fath al-ʕislaamii wajada ʔahl al-maġrib ʔanna-haa fii ʔaaja ʔilaa luġa
 yafhamuuna bi-haa kullu-hum, wa ʔariiqa yaktubuuna bi-haa maa yuriiduuna
 kitaabata-haa. wa lammaa kaanat al-ʕarabiyya hiya luġat al-ʕislaam wa al-
 qurʔaan fa-qad badaʔuu yuqbiluuna ʕalay-haa wa yataʕallamuuna-haa.

“At the time of the Islamic conquest the people of the Maghreb found that they were in need of a language they would all understand and a way to write whatever they wanted to write. Since Arabic was the language of Islam and of the Koran they began to dedicate themselves to learning it.”

ʕabd al-munʕim sayyid ʕabd al-ʕaal
 (1968) lahjat ʕamaal al-maġrib. tiʔwaan wa maa ʔawla-haa
 [The dialect of Northern Morocco. Tetouan and the surrounding area], p. 38.

Chapter 10

Evaluation of the MSA

This chapter begins with a summary of the patterns of MA/Dutch CS that have been discussed in the previous chapters. Then I will evaluate the Monolingual Structure Approach in the light of these findings: to what extent does the insertion paradigm do justice to the grammatical regularities of CS behaviour? Here the performance of the MSA will be compared to that of the model which has prevailed since the early 1990s: Myers-Scotton's Matrix Language Frame model. In section 3, I will discuss the issue of a 'codeswitching grammar': do the regularities in CS result from the combination of two grammars, or is there a third, independent set of rules for the CS variety? In this section I will review the central idea underlying the insertional approach to CS, namely, that speech production is modular, and modules from two languages can be combined to form a single utterance. The 'third grammar', that is, a distinct set of rules for the CS variety, is presented as a possible outcome of an independent diachronic development of a CS variety. Section 4 addresses the fact that CS displays regular patterns. Insofar as it is supported by the data, the insertional model of the MSA succeeds to a large extent in constraining the possible forms of intrasentential CS. However, real CS behaviour is even more constrained than this, since not all logically possible insertion types actually occur as the description of MA/Dutch in the MSA framework demonstrates. This raises the question of why certain patterns occur whereas others do not. Answering this question constitutes another level of investigation and explanation. Section 4 presents an overview of the questions that emerge following the application of the MSA. These questions will be taken up in Chapter 11, which offers an explorative discussion of several explanatory concepts.

10.1 Summary of the corpus description

Parts of the data description have been summarised at the end of Chapters 5 to 8, and at the end of each section in Chapter 9. For a quick overview of the main facts the reader may refer to the table on page 334. Here I will give a concise recapitulation of the corpus description as a whole, concentrating on the descriptive findings. This summary maintains the organisation found in Chapters 5 to 9, dividing the data into the embedded word and constituent categories in the following order: nouns, adjectives, nominal constituents; verbs; prepositional, adverbial, clausal constituents, and discourse markers. For each category, insertions from both languages will be compared.

10.1.1 Embedded nouns

The MA/Dutch CS is primarily characterised by the insertion of Dutch content words in MA matrices. All respondents insert Dutch nouns and this is by far the most frequent and widespread CS pattern. Some respondents usually assign MA feminine gender to inanimate Dutch nouns, but in general gender assignment is not very consistent. Embedded nouns receive minimal MA inflection: inflection for plurality is realised in Dutch, that is, Dutch plural nouns are inserted when the MA matrix requires the encoding of plurality. Dutch nouns are not subject to morphological derivation. The only MA suffix that combines with Dutch words at all is the definite prefix *l-*, and there is a strong tendency for even this prefix to be omitted before Dutch nouns. Omission was found to be almost total in contexts where it is redundant, notably after demonstrative determiners and the indefinite article *wahed*. On the other hand, sometimes *l-* does surface before Dutch nouns when it constitutes a phonological word with a preceding determiner or preposition. The distribution of Dutch nouns in MA clauses was found to be largely analogous to the distribution of MA nouns: they occur in all syntactic positions and semantic roles. One feature was noted with regard to this: the MA preposition *l* behaves like the definite prefix *l-* in that it tends to be left out before Dutch nominal constituents, which most often consist of just a Dutch noun. Dutch nouns differ somewhat from MA nouns with respect to their modification in MA nominal constituents. Setting aside the often omitted definite prefix, Dutch nouns freely combine with MA non-affixal determiners, possessive, adjunct or complement PPs, and relative clauses. Yet one rarely finds a MA attributive adjective modifying an embedded noun. Instead Dutch attributive adjectives recurrently accompany embedded nouns. Embedded adjective-noun combinations retain the Dutch word order (Adj-N), which is opposite to the MA order. Numerals and other quantifiers display a weaker form of the same tendency: MA and Dutch numerals and quantifiers occur in almost equal measure as modifiers of embedded nouns.

While embedded Dutch nouns of all kinds are abundant in all conversations, there are rather few instances of MA nouns in Dutch matrix clauses; therefore any generalisations must be viewed with some reserve in this case. The attested insertions are of two kinds. One group consists of highly specialized vocabulary referring to typically Moroccan or Islamic concepts. The other consists of a number of MA nouns and adjectives that are embedded when repeated: the word recurs in a Dutch context after it has been used in a MA syntactic context. Insertion that results from repetition does not necessarily involve highly specialised vocabulary. The Dutch matrix in which the repeated MA word is embedded is often restricted to a nominal or prepositional constituent which is itself part of a MA clause. Thus, layered embedding plays a significant role in connection with repetition. We noticed a constraint on the modification of embedded Dutch nouns by means of ML attributive adjectives; in the case of MA nouns in Dutch matrices no such constraint appears to exist.

10.1.2 Embedded adjectives

The insertion of Dutch attributive adjectives is very rare. Dutch adjectives, however, are common as predicates of MA copula constructions. Just as embedded Dutch nouns may only be modified by a Dutch attributive adjective, Dutch predicative adjectives are modified by a Dutch degree adverb, if any. Indeed, embedded predicative adjectives can usually be classified as EL constituents.

There are only two instances of embedded MA adjectives (one attributive and one predicative), but, in view of the small number of MA content word insertions generally, their scarcity is not as conspicuous as that of Dutch attributive adjectives. The attested instances result from the same repetition phenomenon that accounts for a number of the embedded MA nouns.

10.1.3 Embedded nominal constituents

The insertion of Dutch nominal constituents did not turn out to be a very productive CS strategy. In particular, NPs that contain an article or demonstrative determiner along with the head noun are seldom inserted, although they are a common and frequently occurring type of NP in (monolingual) Dutch. The embedded NPs that do occur are nouns determined by a possessive pronoun or a quantifier, and some indefinite pronouns. In addition, the insertion of NPs is rather unconstrained in the predicate position of MA copula constructions. The idiosyncratic CS variety of one corpus respondent is conspicuously more ‘liberal’ when it comes to inserting Dutch NPs, including those that contain articles.

One pattern very characteristic of CS with Dutch as the ML is the insertion of MA nominal constituents in the first position of the Dutch clause. Such clause-initial NPs are not connected with culturally specific vocabulary; instead they are often indefinite pronouns. This phenomenon is most conspicuous when the Dutch matrix is a copula clause.

10.1.4 Embedded verbs

The insertion of Dutch verbs involves a periphrastic construction with the MA verb *dar* “to make, do” and a Dutch infinitive. As a dummy verb *dar* carries all verbal inflection, including object pronouns in the case of transitive embedded verbs. This builds on a construction with *dar* and a verbal noun that exists in MA, even though it is not a frequent feature of the monolingual language. Dutch infinitives can be said to be inserted in this particular MA periphrastic construction. Embedding verbs in such periphrastic constructions is a characteristic of CS with many language pairs. In the case of MA this phenomenon contrasts with the practice of incorporating French and Spanish by attaching MA affixes to the foreign infinitive or verb stem. The periphrastic *do*-construction was found to be very frequent with some respondents, occasionally used by others, and totally absent from the contributions of yet other speakers. The *dar* construction as used by different respondents shows

varying degrees of grammaticalisation, and this is most evident from the different ways of encoding pronominal objects. Some further observations were made regarding the complementation patterns of embedded verbs: pronominal complements are always realised in MA as suffixes of the dummy verb, while lexical Direct Objects and prepositional complements are nearly always realised in Dutch. Such lexical (i.e. non-pronominal) complements of embedded verbs are not always well-formed Dutch constituents: Dutch determiners are often missing in NPs and pronouns are often lacking in embedded PPs. So embedded verbs do not ‘project’ well-formed complement constituents in Dutch. The apparent relationship between embedded verbs and their complement preposition or noun is therefore best conceived of as a collocational link between lexical items. Yet it goes further than mere co-occurrence: the relative order of the embedded Dutch verb and its embedded Dutch complement generally obeys Dutch syntax (the utterances of one respondent form an exception). The same applies to the Dutch adverbs that regularly accompany embedded verbs. Embedded reflexive verbs are sometimes accompanied by the Dutch reflexive pronoun.

Concerning verb-insertion with Dutch as the ML, absolutely no MA verbs were inserted in Dutch clauses.

10.1.5 Embedded prepositional constituents

Embedded Dutch prepositional constituents are fairly evenly distributed among the respondents. Unlike embedded Dutch NPs, embedded PPs are quite similar to the PP constituents in monolingual Dutch. Setting aside the Dutch PPs that occur as complements of embedded verbs, the large majority of the embedded PPs in MA clauses are adjuncts, not complements.

There are only eight examples of MA PPs in Dutch clauses. These are all well-formed MA constituents too. Some of them can be regarded as complements of Dutch verbs; in these cases, however, the selection of the MA prepositions does not seem to be determined by Dutch subcategorisation patterns.

Singly embedded prepositions were not found in the data, although this may be viewed as an artefact of the interpretation and classification of the data within the framework of the MSA. There are some instances of MA prepositions, notably *dyaal* ‘of’ used in the analytic genitive construction, occurring as the single MA lexical element in a Dutch clause. In these cases the MA preposition and its Dutch nominal complement were regarded as constituting an embedded MA PP constituent. This manner of classification was justified by pointing out that the Dutch complements in these PPs are often lacking determiners that would be obligatory in Dutch. Hence the complements of these MA prepositions are similar to the Dutch content words that are inserted in MA clauses.

10.1.6 Embedded adverbial constituents

The insertion of adverbs is the domain of much individual variation among the respondents of the Nijmegen corpus. One respondent frequently used Dutch manner adverbs, whereas another inserted many Dutch modal adverbs in her MA clauses. Embedded adverbs can all be classified as adverbial constituents. Dutch adverbs do not occur as modifiers of MA adjectives or adverbs, and embedded Dutch adverbs are not themselves modified by a MA (degree) adverb. This parallels the restriction on modifying relationships that was observed to hold between embedded Dutch nouns and MA attributive adjectives.

MA adverbs in Dutch matrix clauses are largely limited to place and time adverbs that occur as Topic constituents in Dutch clauses, much like the majority of embedded Arabic NPs.

10.1.7 Embedded clauses

The insertion of Dutch subordinate clauses as constituents in MA matrix clauses and vice versa occurred in the data of some respondents who also engage in intersentential CS. The asymmetry between MA and Dutch as matrix languages does not apply at this level: it is almost equally common to find Dutch clauses embedded in MA matrices as MA clauses in Dutch matrices. A relatively common pattern consists of a MA conditional clause followed by a main clause in Dutch. Embedded clauses are usually well-formed EL constituents.

10.1.8 Discourse marking

Discourse marking was not examined in detail because much of it falls outside the domain of insertional CS. It is noteworthy, however, that aspects of discourse marking can evolve in a language different from that of the matrix clause. The habit of one respondent of using Dutch modal adverbs in MA clauses has already been mentioned. Other Dutch discourse markers that occur with MA clauses are the words for “yes” and “no” which function as turn-taking devices and mark concession or partial agreement with what was said in the preceding turn, and the Dutch question tag *hè* that marks constituents and clauses in order to solicit attention and agreement from the audience. MA discourse markers that occur with Dutch clauses seem to be for the most part clause- or ‘sequence’-initial: these include the emphatic independent pronouns marking Topic shift, the question marker *was*, and a number of conjunctions, notably adversative and causal ones.

10.2 Evaluation of the Monolingual Structure Approach

Before I evaluate to what extent the data of the Nijmegen MA/Dutch corpus corroborate the predictions of the Monolingual Structure Approach, it will be useful

to recapitulate the main features of my proposal. The Monolingual Structure Approach assigns each individual syntactic structure to the grammar of a single language, the ML, and describes codeswitching as the insertion of elements from another language, the EL, into this ML structure. In Chapter 2, two structural levels of matrix structure were identified: the finite clause and the constituent. The ML for the finite clause is the language of the inflection bearing element of the tensed verb. In an ML finite clause, EL constituents may be inserted; these are well-formed constituents according to EL grammar. Constituents such as NP, PP, may themselves also be matrices in which smaller constituents or morphemes ('ultimate constituents') are inserted. On this level, no such unambiguous criterion as 'the finite verb' is available, and the ML is inferred from the internal make-up of the constituent, notably the distribution of function and content morphemes. Of course, a finite clause may function as a constituent within another finite clause, in which case the finite verb will determine the ML of the main and subordinate clause independently from each other. The principle of generalisation, which aims to describe a set of data with the least possible number of insertion types, is a major guideline for the recognition of the ML. Because the ML is determined independently on more than one level of a hierarchic syntactic structure, there can be an insertion inside a matrix structure which is itself embedded in a higher-order structure. I call this Layered Embedding.

The Monolingual Structure Approach does not preclude the insertion of function morphemes in ML constituents. For this reason, the exact delimitation of the categories of function and content morphemes is not critical for my approach. The MSA does predict that embedded function morphemes, like content morphemes and (more complex) constituents, follow the distribution of a corresponding ML category. In the case of function morphemes this means that they are the expression of ML grammatical features. It is only principles of generalisation that lead to the observation that the insertion of single function morphemes in mixed constituents is highly uncommon. More frequent is the insertion of inflected or marked EL content words, where the EL marking represents an ML grammatical category.

The MSA predicts that the distribution of the direct constituent parts of the matrix structure can be ascribed to a single language, to be called the ML. In the case of the finite clause, the direct constituents are first and foremost the verb and its arguments and secondly adjunct constituents; in the case of a matrix constituent, the constituent parts are either lower order constituents or morphemes ('ultimate constituents'). The distribution of a morpheme or more complex constituent concerns both its mere occurrence and its order relative to the other parts of the matrix structure. The predictions of the MSA are proven false when the distribution of a word or morpheme cannot be ascribed to ML syntax, whether this concerns the ML on the constituent or the finite clause level.

To what extent do the data support this approach to codeswitching? A range of recurrent CS patterns have been recognised, including both patterns that fit in neatly with the insertion paradigm and patterns that do not. Hereafter I will review the performance of the MSA, starting with the CS patterns that support this framework

(2.1) and subsequently discussing the challenges (2.2). Section 2.3 addresses not so much counter-examples to the MSA as its limitations. Recapitulating the main points of Chapter 3, I will demonstrate the inadequacy of insertional approaches like the MSA when it comes to modelling the distribution of discourse markers in bilingual texts. As the MSA is a new proposal, it will be useful to compare its performance to that of the established model in the insertion framework, viz. Myers-Scotton's Matrix Language Frame model.

In both proposals the concept of insertion is similar but, crucially, the definition of the matrix language diverges widely. Remember that in Myers-Scotton (1997) the ML is defined in three ways: [1] "the language projecting the morphosyntactic frame for the entire CP which shows intrasentential CS", [2] "the language contributing more morphemes in a sample of discourse-relevant intrasentential CS (minimally two contiguous CPs, either from a single speaker or from an adjacency pair produced by two speakers)", and [3] "the language of more morphemes in the discourse as a whole, including monolingual stretches" (Myers-Scotton, 1997: 23). Remember also that the first criterion "is operationalized as the *morpheme order* and *system morpheme* principles of the MLF model" (Myers-Scotton & Jake, 1995: 83). (See Chapter 1, p. 37 ff. for a discussion of these criteria.)

The definition of the ML has consequences for the way in which the insertion of constituents ('EL islands' in Myers-Scotton's terminology) is constrained. In the MSA, an EL constituent cannot include the inflection of the finite verb as this defines the ML, while constituents like IP or I' (I-bar) are possible in the MLF model. As a consequence, unlike the MLF model, the MSA claims that the ML predicts the distribution (presence and order) of the constituents inside the finite clause. "Some aspects of EL islands," Jake & Myers-Scotton (1997: 6) write, "may be determined by the ML, for example, their position in the larger CP". The crucial word here is 'may'. While the MLF model allows types of EL constituents that include the finite verb, the MSA recognises the possibility of single morpheme constituents. Indefinite pronouns, for instance, are classified as types of NP by virtue of their distributional properties. Myers-Scotton's MLF model predicts that only EL content morphemes occur in ML + EL constituents (the System Morpheme Principle), while EL system morphemes may occur in EL islands. If the MLF model were to allow for single morpheme EL islands, this would constitute an escape hatch to the System Morpheme Principle. In order to avoid this, Myers-Scotton (1993b: 138) explicitly stipulates that "all islands must be composed of at least two lexemes/morphemes in a hierarchical relationship".

10.2.1 Supportive evidence

The data at large corroborate the proposed definition of the ML on the finite clause level and confirm the possibility of a different, independent ML on the constituent level.

10.2.1.1 The definition of the ML on the finite clause level

If the ML is to determine the distribution of constituents in the finite clause, its definition needs to refer to a feature of the finite clause itself, rather than to factors outside the clause such as larger stretches of discourse or sociolinguistic considerations. The following extract from the MA/Dutch text corpus will clarify this. Within the predominantly MA passage in (1), we notice the Dutch string *ben ik mezelf* ‘I’m being myself’ in the second line. In order to account for the Verb-Subject word order within this Dutch string, we need to assume that it is part of a Dutch clause which also includes the preceding MA PP *mša-k* ‘with you’. In the Dutch finite clause *mša-k ben ik mezelf*, the topical position of the non-Subject constituent *mša-k* triggers the inversion of Subject and finite verb, in accordance with the ‘verb second’ rule in Dutch. Consequently, *mša-k* is analysed as an EL constituent within this matrix clause.

- (1) ana-ya ma ka-n-dir-š *mezelf voorstell-en zoals ik echt ben*,
 1SG-EMPH NEG ASP-1-do-NEG myself present-INF like I really am
- fhem-t-li-ya? ana-ya, ana-ya, mša-k *ben ik mezelf*. mša-k
 understand-2M-to-1SG 1SG-EMPH 1SG-EMPH with-2SG am I myself with-2SG
- kūn-t *mezelf*, b š-šəḥḥ mša bba u mša mma u mša xu-ya
 be-1SG myself with DEF-reality with father and with mother and with brother-1SG
- xu-ya waxxa - f l-mustaqbel llaḥu šalem, ġadi n-kun *mezelf*,
 brother-1SG alright in DEF-future God know·PART FUT 1-be myself
- b š-šəḥḥ durka šawda ana-ya *iemand anders*, ya-k?
 with DEF-reality now still 1SG-EMPH someone different QTAG-2SG
 ‘I don’t show myself the way I really am, you see? I er, with YOU I’m being myself. With you I was being myself, but with my father and my mother, and with my brother - my brother ok - well, maybe in the future God knows, I will be myself. But now I am still being someone else, you see.’

The above analysis is consistent with the principle of considering ‘the inflection of the finite verb’ as the criterion which identifies the ML. As earlier mentioned, the 1997 version of Myers-Scotton’s Matrix Language Frame (MLF) model defines the ML by means of three criteria. Let us start with her criterion [2] ‘the language contributing more morphemes in a sample of discourse-relevant intrasentential CS (minimally two contiguous CPs, either from a single speaker or from an adjacency pair produced by two speakers)’. In the above example, the counting of MA and Dutch morphemes in the immediately preceding and following sentences identifies MA as the ML. This is a possible alternative, but a matrix language thus defined does not predict the word order in *mša-k ben ik mezelf*. Besides, if MA is considered as the ML, the classification of the string *ben ik mezelf* poses a problem for the MLF

model. The MLF model might classify *ben ik mezelf* as a Dutch ‘EL island’ or perhaps as a series of EL islands. Since EL islands must be well-formed constituents according to EL grammar (Myers-Scotton, 1993b: 77-8), the first solution requires the string *ben ik mezelf* to be classified as a constituent. In X-bar theory to which the studies in the MLF framework refer, this string might be classified as a Dutch IP (Inflectional Phrase) constituent (cf. Figs. 3.2 and 3.3 in Ch. 3, p. 134). However, the internal word order of this IP can only be explained by referring to a Dutch structure that includes both *ben ik mezelf* and the MA constituent *m ša-k*. On the other hand, note that *ben*, *ik*, and *mezelf* cannot be EL islands because the MLF model does not provide for the possibility of single morpheme islands (1993b: 138). The possibility remains of regarding *ben ik* ‘am I’ and *me-zelf* ‘my-self’ as two contiguous EL islands, each one consisting of two morphemes. The relative order of the three strings *m ša-k*, *ben ik*, and *mezelf* can be accounted for by reference to MA grammar - compare the MA translation *m šak // ka-nkun // raš-i* - but in this case it will be difficult to argue for the constituency of the string *ben ik*.

Then let us try to apply criterion [1]: “the language projecting the morphosyntactic frame for the entire CP which shows intrasentential CS”. This criterion “is operationalized as the *morpheme order* and *system morpheme* principles of the MLF model” (Myers-Scotton & Jake, 1995: 983). However, in the CP *anaya .. mezelf* there is no mixed constituent in the sense of the morpheme order and system morpheme principles, that is, a mixed constituent in which the ML provides the system morphemes and the morpheme order. I will refrain from testing criterion [3] “the language of more morphemes in the discourse as a whole, including monolingual stretches”. Counting morphemes on the discourse level is not a feasible option, if only because the delimitation of ‘the discourse as a whole’ is vague.

In any case, Myers-Scotton’s criteria will designate one ML for the entire CP *anaya .. mezelf* in line 2 of example (1). (Recall from Chapter 3 that the MLF model situates clause-initial discourse-emphatic pronouns like *anaya* in the ‘Spec of CP’ position, that is, within the CP.) If MA is identified as the ML for this CP, it will account for the emphatic pronoun *anaya* but not for the constituent order in *m ša-k ben ik mezelf*. Conversely, if Dutch is to be the ML, the presence of the Arabic pronoun *anaya* is not accounted for, since this is not a common way to indicate Topic shift or contrastive Topics in Dutch.

10.2.1.2 The definition of the ML on the constituent level and layered insertion

The MSA diverges from other matrix language approaches in its explicit assumption of an ML on the constituent level. Just as a conversation, speech turn or utterance or any other ‘discourse-relevant’ sample cannot fully predict the ML on the finite clause level, the finite clause does not necessarily govern the internal structure of each complex constituent within the finite clause. The possibility of having EL constituents in a matrix clause already makes this apparent. Because the EL constituent can itself be a mixed constituent containing elements from two languages, as in the case of the NP *religie dyal-na* ‘our religion’ in the example below, it is essential to acknowledge

and identify an ML on the constituent level which is independent of the ML on the finite clause level.

For the next example (2), the MSA identifies Dutch as the ML on the finite clause level, due to the Dutch finite verb *is* “is”.

- (2) (*ook eh ge.. eh religie bijvoorbeeld:*) *hna, religie dyal-na is veel sterker*
 also er g.. er religion for-example 1PL religion of-1PL is much stronger

eh in ons eh ingeprint, in ons hart, in onze hersens

er in us er drummed in our heart in our brains

“Also er rel.. er religion, for instance. As for us, our religion is much more er drummed into us, in our heart, in our brains.” (Samir, ex. (22) in Ch. 9)

This ML is consistent with the constituent order in the clause. The NP *religie dyal-na* functions as the Subject, as it is immediately followed by the finite verb which agrees with it in person and number. (The MA pronoun *hna* which marks Topic shift is considered to be outside the finite clause; I will return to it presently.) *religie dyal-na* is analysed as a MA NP, that is, MA is the ML for this constituent. This is evidenced by the analytic possessive construction itself, the morpheme order and the presence of MA function morphemes *dyal* and *-na*. With regard to function morphemes, remember that the MSA does not preclude the possibility of their being EL morphemes, although the principle of generalisation, which aims to limit the number of insertion types, leads to an analysis which attributes function morphemes to the ML. However, setting aside the reluctance vis-à-vis the admission of ‘function morpheme insertion’, *dyal* and *-na* cannot be embedded morphemes because they do not fit in a slot in the Dutch constituent: Dutch would use a pronominal possessive pronoun in this context (*onze religie* “our religion”). (True, an analytic possessive with pronominal possessor does occur in Dutch, but it has a marked modal value very much like the English counterpart “that religion of ours”.) Inside the MA constituent *religie dyal-na* we find an embedded Dutch content morpheme: *religie*. Thus at the same time (2) demonstrates the possibility of what I have called layered insertion.

Layered insertion is another feature that distinguishes the MSA from most other insertional approaches to CS, with the exception of Nishimura’s (1986) account of Japanese/English. The possibility of layered insertion is a logical corollary of the recognition of lower and higher order constituents (including morphemes as ‘ultimate constituents’) as elements that can be inserted.

Myers-Scotton’s MLF model does not provide for the possibility of different matrix languages on the finite clause and the constituent levels. If the number of MA and Dutch morphemes in the immediately preceding or following sentences is taken as the criterion, the MLF model would identify Dutch as the ML for the entire passage cited in (2). Then the MLF model classifies MA *dyal-na* as an EL island. However, Dutch as the ML of this text sample cannot account for the occurrence of this possessive PP in this syntactic position. This does not necessarily disqualify the MLF

model since the latter makes no explicit predictions as to the distribution of EL islands. Yet the MSA performs better here: Dutch, as the ML on the finite clause level, accounts for the clause-initial position of the NP *religie dyal-na*, whereas the internal structure of this NP is explained by identifying MA as the ML on the level of the constituent.

Note in passing that if Myers-Scotton's criteria identify Dutch as the ML for the CP *hna ... hersens* in (2), the occurrence of MA *hna* poses another problem for the MLF model. This pronoun would be classified as an EL content morpheme in an ML + EL island. (The classification of *hna* as a 'system morpheme' would falsify the System Morpheme Principle, and the model does not allow for single morpheme EL islands). The Morpheme Order Principle of the MLF model predicts that the ML, in this case Dutch, determines the surface morpheme order in the ML + EL constituent. The model is challenged here because it is hard to conceive of a constituent in which Dutch accounts for the presence, let alone the position of the MA pronoun *hna* in the sentence.

Myers-Scotton's (1997) criteria [1] and [3] may designate MA as the matrix language in (2). In that case the MLF model analyses the string *is veel sterker ... in onze hersens* as an EL island (i.e. EL constituent), or as a series of EL islands. This raises once more the question of the constituency of this string. Again, X-bar theory facilitates the identification of EL islands: according to this theory the IP constituent can be analysed as comprising two immediate constituents: the Subject in the Spec of IP position, and the I' (I-bar) phrase, consisting of the inflection and the VP (see again Figs. 3.2 and 3.3 in Ch. 3, p. 134). Hence the string *is veel sterker ... in onze hersens* may be regarded as an I' island according to the MLF. But here again the word order of this Dutch I' constituent depends on something outside of it, namely whether it is immediately preceded by the Subject NP, as in the present case, or a non-Subject constituent, as in (1).

10.2.1.3 General assessment

The larger part of the data can be adequately described in terms of insertion patterns. Most instances of MA/Dutch CS concern Dutch content nouns embedded in MA nominal constituents which participate in MA clauses. Dutch verbs in MA clauses can be analysed as insertions into the MA periphrastic construction [*dar* VERBAL NOUN]. In addition we have seen smaller numbers of other embedded Dutch content words and of embedded Dutch constituents. Dutch adjectival, prepositional and nominal constituents occur freely as predicates of MA copular constructions.

Dutch word order rules are instrumental to the identification of the matrix language and of EL constituents. In the case of embedded Dutch clauses which constitute a constituent of a MA clause, the 'embedded' status of the Dutch clause is corroborated by its distinct subordinate clause word order. Note the final position of the finite verb *ontmoet* 'meet' in the conditional clause in (3). Where Dutch is the ML on the clause level, Dutch word order demonstrates the embedded status of MA constituents and the ML status of Dutch: embedded MA constituents turn out to occur mainly in Topic

position, and this position is always followed by the finite verb in Dutch main clauses. This Dutch syntactic rule was discussed above with respect to examples (1) and (2). It also applies when a subordinate clause precedes the main clause. Hence it was possible to identify MA subordinate clauses such as conditionals as embedded constituents, cf. (4).

- (3) *als je een meisje ontmoet, eh ġadi t-ħebb-ha?*
 if you a girl meet er FUT 2-love-3F
 “If you meet a girl, will you love her?” (Samir, ex. (8) in Ch. 8)
- (4) *ila der-t-l-ek ana l-ŧerđ u t-hezz-u nta te-ŧti-h-l-u, is dat*
 if do-1SG-for-2SG 1SG DEF-paper and 2-carry-3M 2M 2-give-3M-to-3M is this
duidelijk
 obvious
 “If I write the paper for you and [then] you give it to him, this [i.e. the fraud] will be obvious.” (Hocine, ex. (59) in Ch. 9)

Significantly, the Monolingual Structure Approach made it possible to describe the data corpus in terms of a limited number of more or less recurrent insertion types.

10.2.2 Challenges

Using the principles of the Monolingual Structure Approach I described a number of phenomena that actually challenge the idea of insertion central to this approach. I will discuss these challenges below, and here again, I will examine the interpretation of the same phenomena according to Myers-Scotton’s Matrix Language Frame model.

The following types of counter-examples to the MSA are identified: firstly, certain embedded categories, notably verbs, go together with EL syntactic features inside the matrix structure. Secondly, there is the omission of certain function morphemes in insertion contexts, as well as the occasional ‘superfluous’ morphemes due to marking in both languages. Thirdly, the ML fails to predict the word order of embedded attributive adjectives.¹²

10.2.2.1 Impact of embedded verbs

A major challenge refuting the idea of a matrix language arises from EL words that import aspects of EL grammar into the ML structure. Examples are the Dutch word

¹² In addition, recall that in view of the data on other language pairs it is probable that the ML also fails in predicting the word order properties of embedded modal adverbs and adverbs with discourse sequencing functions (cf. Chapter 3). However, I am unable to verify this for the MA/Dutch data due to insufficient detailed information on the syntax of such adverbs in these languages.

(5) hna der-na, verdorie, — maatschappelijk werker aanstell-en voor die
 1PL do-1PL damn-it [a] social worker appoint-INF for those

marokkanen
Moroccans
‘We, damn it!, appointed a social worker for those Moroccans.’ (Samir)

(6) ma ne-qder-š n-dir *me eigen concentrer-en*, wella ne-qra ši haža
 NEG 1-can-NEG 1-do my own concentrate-INF or 1-read INDEF thing
 “I can’t concentrate or read anything.” (Jamal)

In order to save the MSA, it might be argued that the Dutch infinitive and its complements together constitute an EL verb phrase constituent. After all, in monolingual MA the (verbal) noun in the periphrastic constriction [*dar* + N] may also be accompanied by a prepositional complement: [*dar* + N + PP]. (See examples (3)-(5) in Chapter 6.) This solution is somewhat complicated, however. Firstly and most importantly, the Dutch infinitive plus its complements is not always a well-

formed EL constituent. The PP complement, like *voor die marokkanen* “for those Moroccans” in (5), is usually well-formed according to Dutch grammar, unless the complement of the preposition is a pronoun (cf. (9) hereafter). Lexical Direct Object complements of embedded Dutch verbs, however, are typically without any determiner even if Dutch grammar would require one. The missing *een* “a” in (5) above exemplifies this. Moreover, it is not at all obvious that the missing determiner is a Dutch one in these cases: sometimes we find a MA determiner with a Dutch noun which functions as the Direct Object of an embedded Dutch verb like the demonstrative *hadak* in (7) (cf. Ch. 6, section 3.3). Besides, the MA determiner *l-* tends to be omitted before Dutch nouns in all contexts. Therefore, the Dutch Direct Object of an embedded verb may in general be classified as a MA constituent, rather than a Dutch one. Such an analysis would, moreover, be in line with the insertion of Verb-Object collocations in periphrastic *do*-constructions in other language pairs such as Greek/English (Seaman, 1972), where the embedded DO noun does receive ML determiners, as the following example shows. Likewise in Turkish/Dutch (Backus, 1992, 1996b), where the embedded complement of an embedded verb also receives Turkish case marking. (In these two language pairs, the relative order of verb and complement does not differ for the ML and the EL, however.)

- (7) *daʔimen mnin der-t hadak erving opdoe-n, (..)*
 always when do-1SG DEM experience get-INF
 “Ever since I had this experience, (...)” (Samir, ex. (74) in Ch. 6)
- (8) *káni measure to power*
 he·does measure·INF DEF power
 “He measures the power.” Greek/English (Seaman, 1972: 167)

The MSA can still account for examples like (5) and (7) by arguing that they display layered embedding. According to this analysis, the string *hadak erving opdoen* in (7) would be a Dutch VP which contains the MA NP *hadak erving*. This NP, in turn, forms a matrix in which the content morpheme *erving* is inserted. In this way, both the Dutch DO-Verb constituent order and the MA determiner in the DO constituent can be accounted for.

The second complication concerning the ‘VP insertion’ analysis for MA/Dutch is that, while the infinitive and its complements may constitute a constituent in Dutch, this constituent does not correspond very closely to an ML constituent type, particularly if it contains a reflexive or personal pronoun. In sum, it must be recognised that the embedded verb has an influence on the word order in the MA clause exceeding that of the EL constituent.

The embedded Dutch infinitives and their lexical complements do not pose too many problems for Myers-Scotton’s MLF model, simply because it does not make explicit predictions about the distribution of constituents in the clause, i.e., their presence

and their order relative to each other. If the Dutch complement of an embedded Dutch infinitive is a well-formed Dutch NP or PP constituent, the MLF model might classify the infinitive and its complement together as an EL island, e.g. *me eigen concentreren* in (6). If the complement is not a well-formed EL constituent, the MLF model can account for both the infinitive and the lexical complement as being EL content morphemes, each of them embedded in its own ML + EL constituent. In example (7), for instance, the noun *ervaring* is inserted in the mixed nominal constituent *hadak ervaring*, and the verb *opdoen* is embedded in another mixed constituent, viz. one of the X-bar type constituents VP, I', IP or CP. This is a possible analysis, but note that the relative order of *hadak ervaring* and *opdoen*, which must be attributed to Dutch, remains unaccounted for in the MLF model. A problem for the MLF model arises when both *ervaring* and *opdoen* are analysed as being part of the same ML + EL constituent (VP, I', IP or CP): firstly, the model explicitly allows for SINGLY embedded EL content morphemes only, see the formulation of the Morpheme Order Principle, cited on p. 36. Secondly, this principle predicts that “surface morpheme order will be that of the ML”.

On a more general level Myers-Scotton acknowledges that the periphrastic *do*-constructions common in many CS varieties, though not counter-examples to her model, qualify the conception of CS as being the insertion of content morphemes and constituents. Therefore, Jake & Myers-Scotton (1997) assume periphrastic *do*-constructions under a series of phenomena called “compromise strategies”.

Concerning the MSA, it must be conceded that examples like (5) and (6) invalidate the conception of the ML clause as the structure in which all constituents are inserted. On the other hand, such patterns do confirm that the verbal inflection is the last foothold of the ML on the finite clause level. Clauses like the ones in (5) and (6) still display a great deal of MA grammar, precisely in the expression of such verbal categories as tense, mood, aspect and modality. The weighty impact of embedded verbs is attested only in the case of those respondents who are fluent in Dutch, while the insertion of verbs in the *dar* + infinitive construction itself is also attested in the speech variety of (beginning) learners of Dutch. The CS variety in which embedded Dutch verbs trigger Dutch complements and Dutch syntactic rules within the MA finite clause may be the outcome of a diachronic development of a CS variety that is accounted for in the MSA. Such a diachronic development is conceivable either on the individual level or on the level of smaller speech communities. I will expand on this point in section 3.2 below. Following the diachronic process in which embedded Dutch verbs trigger more and more Dutch lexical and grammatical elements, I speculate that at a certain point the selection of a Dutch verb also triggers the realisation of Dutch inflectional categories associated with it. As soon as this happens, the MSA will classify the finite clause as Dutch. I call this the Matrix Language Turn-over, which will be discussed in section 2.1 of Chapter 11.

10.2.2.2 Bare forms and double marking

Some minor challenges to the MSA come from the occurrence of so-called bare forms, that is the recurrent omission of certain function morphemes, and from instances of double marking whereby a grammatical category is marked in both the ML and the EL. In the case of MA/Dutch CS various types of bare forms have been discussed in the chapters on insertion with MA as the matrix language. In principle it cannot be established whether a MA (ML) morpheme is left out, or a Dutch (EL) one. If the function morpheme does surface in a number of the instances, we might argue that the missing morphemes are from the same language as their surfacing counterparts. But it could equally well be maintained that, if the realised function morphemes come from one language, the missing ones must be from the other language. The concept of matrix language implies that the ML determines the presence or absence of function morphemes. Therefore, the MSA tends to assign missing morphemes to the matrix language.

In Chapter 5 (section 1.5) I discussed the recurrent omission of the MA definite prefix *l-* before Dutch nouns. Here, it is plausible that the missing morpheme is the MA one because in the monolingual language the definite prefix *l-* also fails to surface before various words of Berber and Romance origin. In section 6 of Chapter 5 it was shown that the MA preposition *l* tends to be omitted before Dutch nouns as well. The pronominal complements of embedded Dutch verbs either emerge as pronominal object suffixes on the MA auxiliary verb *dar* or they are omitted (Ch. 6, section 2.3). In the case of embedded Dutch PPs the omission of pronominal complements seems to be the general rule (Ch. 6, section 2.6). An example is reproduced here in (9).

- (9) ka-y-dir-u ___-in trapp-en
 ASP-3-do-PL [it]-in step-INF
 “They step in [it].” (expression for “they let themselves be fooled”)
 (Samir, example (41) in Ch. 6)

Also, when Dutch reflexive verbs are inserted, the Dutch reflexive pronoun is usually omitted (Ch. 6, section 2.7). In such cases we can safely assume that a Dutch element is left out rather than a MA one, since most examples concern Dutch inherently reflexive verbs that do not correspond to reflexive verbs in MA.

Instances of double marking are less frequent in MA/Dutch. Occasional examples of double Object pronouns occur with embedded Dutch verbs in Samir’s data; see Chapter 6, section 2.5 from which example (10) below is reproduced.

- (10) ka-t-dir-li-ya mij aanwijz-en
 ASP-3F-to-1SG me point-out-INF
 “She pointed at me.” (Samir, example (37) in Chapter 6)

An isolated instance of double marking is found in example (96) of Chapter 6, where subordination is marked both in MA by [*baš* + subjunctive] and by the Dutch particle

te. As double marking is not a recurrent feature of MA/Dutch CS, I will not press this point here.

The ‘missing’ morphemes in bare forms are counter-examples to the MSA because according to this approach the ML determines the distribution of morphemes, and distribution is taken to comprise both the presence or absence of morphemes and their relative order. With the Matrix Language Frame model, on the other hand, ‘missing’ morphemes do not constitute outright counter-examples because this model makes no predictions with respect to the presence or absence of ML morphemes. The predictions of the MLF model are restricted to the morpheme order and the inhibition of EL system morphemes. However, Myers-Scotton and her co-authors appear to admit that bare forms are at odds with the spirit of the insertional approach to CS: bare forms are presented as another type of “compromise strategy” (Jake & Myers-Scotton, 1997).

Double marking is more problematic for the MLF model when it involves the occurrence of EL ‘system morphemes’ in mixed constituents; see the discussion of double plural marking in Swahili/English (p.36). In a case like (10) above, the problem can be solved for the MLF model if the string *mij aanwijken* is classified as an EL island.

2.2.3 Attributive adjectives

The MA/Dutch data show that the ML on the constituent level does not determine the word order of embedded attributive adjectives. Very few attributive adjectives were embedded at all, but the only MA adjective forming part of a Dutch constituent displays the MA postnominal order rather than the Dutch (ML) prenominal order. This example, discussed in Chapter 9, is reproduced here: the MA adjective *mwessxa* “dirty” is part of the Dutch NP *een straat mwessxa*.

- (11) ma ta-ye-bġi-w-š ɖyur mwessx-in langs een straat mwessx-a
 NEG ASP-3-want-PL-NEG house-PL dirty-PL along a street dirty-F
 “They don’t want dirty houses along a dirty street.”
 (Hayat, ex. (9) in Ch. 9)

Clearly the ML on the constituent level does not impose ML word order on the attributive adjective. At most, the ML as defined in the MSA can be one of several factors determining the noun-adjective order. Further research is needed to clarify this, however.

Embedded attributive adjectives that display source language word order are equally troublesome for the Matrix Language Frame model, as they defy the Morpheme Order Principle. The above example is difficult to interpret within this model because it involves layered embedding. True, the MLF model might consider the string *langs een straat* as a Dutch EL island (on the assumption that the model identifies MA as the ML), but this disregards the position of MA *mwessxa* as a

modifier of *straat*. Even if Dutch were identified as the ML for the ‘discourse-relevant sample’ in which (11) occurs, the analysis of this example would remain problematic for the MLF model. If the adjective *mwessxa* were an EL content morpheme, it would defy the Morpheme Order Principle. Alternatively, the adjective (as well as the MA string *ma .. mwessxin*) could be analysed as an EL island.¹³ This would partially save the MLF model, while leaving the surface word order unaccounted for.

10.2.3 Discourse markers

Forms of extra-clausal discourse marking do not constitute counter-examples to the MSA since these are exempted from the applicability of this framework. This was discussed in more detail in Chapter 3, but it is worthwhile to demonstrate once more why these phenomena cannot be dealt with in an insertion model in the light of the MA/Dutch corpus description. For this purpose, consider two discourse marking devices that are much used in the MA/Dutch CS variety: the MA clause-initial emphatic pronouns which signal Topic shift and contrastive Topics (cf. *hna* in (2)), and the Dutch question tag *hè* which occurs clause-finally and after major constituents (Ch. 8, section 2.3). These two markers may be taken to represent clause-initial and clause-final discourse markers in general.

- (12) *šhal men xetṛa y-ḍerḇ-u u bḃa-h ma ma y-dir-l-u walu hè*
 how-many of time 3-hit-3M and father-M NEG NEG 3-do-to-3M anything QTAG
 ‘He hits him so many times, and his father can’t do anything to him, you see?’
 (Younes, from ex. (27) in Ch. 8)

We can conceive of a matrix structure that encompasses both the finite clause and the discourse marker. In order to integrate these phenomena in the insertion approach to CS we have to define the ML for these encompassing structures and subsequently either of its immediate constituents can be recognised as an EL constituent. The ML has to be defined such that it predicts the distribution of clause-initial Arabic Topic pronouns and of the Dutch question tag *hè*, respectively. On the other hand, it is the discourse marker rather than the ‘finite clause’ constituent which is language-specific. If we study these discourse marking devices we will come to the conclusion that it is Moroccan Arabic that determines the use of Topic pronouns while Dutch determines the distribution of *hè*. In other words, the distribution of these discourse marking devices is generally best predicted by their own source language (Chapter 3).

Hence, if the ML on the supra-clausal level is to predict the distribution of the discourse marker, this ML must be defined as the language of the discourse marker

¹³ There seems to be nothing in the MLF model that prevents the interpretation of *mwessx-a* as an EL island, the only requirement being that “all islands must be composed of at least two lexemes/morphemes in a hierarchical relationship” (Myers-Scotton, 1993b: 138).

itself. Of course we can follow this procedure in order to subsume these phenomena in the MSA, but a matrix language thus defined does not imply any meaningful generalisation. Indeed, it amounts to a roundabout way of saying that the distribution of the discourse marker is determined by its source language. There is no other independent criterion that predicts the use of this discourse marker and its order relative to the finite clause constituent.

Defining the ML for a larger stretch of text is certainly no solution. If we were to assume one ML ‘on the discourse level’ for an entire speech turn, a narrative or a conversation, the model would immediately run into trouble when discourse markers from both languages would be used. There is no reason, for instance, why a finite clause in either MA or Dutch could not at the same time be preceded by a MA discourse-emphatic pronoun and followed by the Dutch question tag *hè*. Hence the Matrix Language Frame model will also fail as long as the ML is determined on the basis of the number of morphemes in a discourse-relevant sample. Setting aside the problem of defining such a sample, I see no reason why a MA Topic pronoun could not occur in a discourse relevant sample in which the majority of morphemes is Dutch or, conversely, why the Dutch question tag *hè* could not occur in a sample with a majority of MA morphemes.

10.2.4 Summary

In general the ML as defined in the MSA makes the correct predictions for the distribution of morphemes in a mixed constituent, and the distribution of constituents in the finite clause. Due to the possibility of layered insertion which is recognised by the MSA, this approach is successful in accounting for instances of CS which are difficult to deal with in most other insertional models of CS. I have shown that approaches which define the ML for a larger stretch of discourse, like an utterance, a conversation or the CS variety as a whole will not be able to account for the distribution of constituents and morphemes in the MA/Dutch text corpus.

On the other hand, it was shown that the MSA fails to predict the word order of attributive adjectives and the omission of certain function morphemes in MA/Dutch. These complications were foreseen and discussed in the outline of the MSA in Chapter 2. Another complication is that in some idiosyncratic varieties of MA/Dutch, the insertion of Dutch infinitives and their complements in the periphrastic construction with *dar* ‘to do’ leads to configurations which can no longer be fully interpreted in an insertion paradigm. The counter-examples to the MSA are not better accounted for by any other insertional approach to CS. That is, a different definition of the matrix language will not help in interpreting these phenomena in terms of insertions.

Finally, I stressed that the distribution of clause-external discourse markers cannot be predicted in an insertion model of CS. The best predictor for their distribution appears to be their source language. No other independent criterion can be found to

define the ML on the supra-clausal level such that it will consistently predict the presence of clause-external discourse markers from either language.

10.3 How many grammars are involved in CS?

Perhaps the most fundamental discussion in CS studies revolves around the question of whether the regularities in codeswitching behaviour result from the combination of two grammars or from a 'third grammar' specific to the CS variety, where this third grammar may subsume (parts of) the two monolingual grammars. Giesbers (1989: 46-50) presents an overview of this discussion. The Monolingual Structure Approach is clearly based on the former assumption: it is inspired by the idea that the production of an utterance involves a number of modules or components, and that CS results from the combination of modules from two languages. Now that the Nijmegen corpus of MA/Dutch has been explored in the framework of the MSA it becomes apparent that the data corroborate the insertion approach only to a certain extent. The frequent insertion of Dutch infinitive verbs in the periphrastic construction with MA *dar* and the far-reaching syntactic consequences this has in the speech variety of some respondents creates the impression that we are, after all, witnessing the emergence of a new speech variety with its own set of rules. This section reviews the concepts of 'modularity' and 'third grammar' in the light of the MA/Dutch data.

10.3.1 Modularity

The insertional approach to CS implies modularity. It has as basis the idea that the production of an utterance is organised in more or less independent components such as the content word lexicon, components which assemble lower and higher order constituents, and various aspects of discourse marking. Further components need to be added to this list, in particular components that relate to the articulatory system. However, pronunciation and prosody have not been investigated in this study and have thus far received scant attention in CS studies generally. Therefore I will confine myself to discussing the morpho-syntactic aspects of mixed utterances.

The independence of these components from each other is substantiated insofar as each component can evolve in a language different from that of the other components. Independence is even observable in the case of components which are in a hierarchical relationship to each other, that is, when the output of one component is the input to the other. The fact that ultimate or more complex constituents from two languages can combine to form a higher order structure in either language further indicates that at least certain components can be defined crosslinguistically. Or, to put it differently, this shows that bilingual speakers 'observe' a crosslinguistic congruence of components.

Hence speakers can draw freely on the content word lexicon of one language while producing a constituent according to the grammar of another language (the ML on constituent level); they can also produce a nominal, prepositional or adverbial

constituent in one language and insert it as a constituent in a higher order structure that is assembled according to the grammar of the other language (the ML on clause level). Also, discourse marking devices such as question tags can be taken from one language while the rest of the utterance is in the other. The concept of modularity is best illustrated by the insertion of EL constituents. The EL constituent is the output of an EL syntactic component and forms the input to a higher-order component of the ML which assembles the matrix constituent or clause.

10.3.2 A grammar for the CS variety

The insertion paradigm assumes that monolingual elements that make up the mixed finite clauses represent the monolingual language varieties. If these monolingual parts of the CS variety should turn out to be different from what is found in the monolingual variety, we could then consider recognising the CS variety as distinct, possessing its own distinct grammar. The question is: how much difference is needed for the CS variety to acquire its autonomy as a distinct set of rules?

The monolingual elements in the CS variety will always differ from the monolingual speech variety with respect to the relative frequency of different word classes. The Dutch words embedded in MA/Dutch, for instance, are not representative of the monolingual language as a whole. Likewise, the periphrastic construction *dar* plus verbal noun is occasionally found in monolingual MA, whereas it is pervasively used with Dutch infinitives in the CS varieties of some speakers. In some CS varieties, like those of the Hamadi brothers, the *dar* construction is grammaticalised and serves to embed almost any kind of Dutch infinitive and its complements. In these speech varieties the construction diverges qualitatively from its counterpart in monolingual MA. A striking symptom of this grammaticalisation is the fact that in Samir's variety pronominal Direct Object complements of the embedded infinitive are usually marked as Direct Object suffixes on *dar*. Thus in at least some idiolectal CS varieties the monolingual MA component differs from monolingual MA and even from the same individual speaker's variety of MA. Whether this is sufficient reason to speak of a 'third grammar' for the CS variety is ultimately a matter of subjective judgment (cf. Myers-Scotton, 1993b: 151).

Even if one decides to regard the CS variety as possessing a 'third grammar', most of its grammatical rules will be efficiently described by reference to its monolingual 'parents'. Indeed, the insertion approach followed in the MSA adequately describes such a variety, firstly by classifying the types of insertions that occur in either matrix language and secondly by pointing out the phenomena which do not fit well into the insertion paradigm.

The degree to which the CS variety diverges from what can be described as the combination of elements from two languages, is the outcome of an independent diachronic change of a CS variety, which in its initial shape was nothing more than the combination of elements from two languages. The diachronic development of a CS variety can be conceived of both on the idiolectal level and on the level of

smaller and larger speech communities. On the one hand, individual speakers may change their linguistic behaviour over time independently of other speakers; on the other hand, members of a speech community will imitate each other's linguistic behaviour to a certain extent. The fact that speakers tend to follow each other's behaviour can lead to community level conventionalisation of CS patterns, while at the same time idiolectal changes can lead to community level change so that the in-group variety evolves as a whole.

The individual variation found in the Nijmegen corpus illustrates this process. Although we do not actually have diachronic data, we may assume that the idiolectal CS varieties which very closely resemble monolingual MA and contain very few insertions of Dutch elements represent earlier stages of a possible diachronic development. After all, in the history of Moroccan migration to the Netherlands, monolingualism in MA predates MA/Dutch bilingualism and hence, mixed speech varieties which include few Dutch elements, both quantitatively and qualitatively, precede mixed speech varieties containing many Dutch elements (cf. section 4 in Ch. 9). The respondents Mustafa and Warda represent such 'early' varieties in the Nijmegen corpus. As learners of Dutch, their confrontation with the bilingual speech mode is relatively recent and thus they are less likely to have acquired existing conventionalised patterns of bilingual speech behaviour or to have developed such routines on their own (although neither possibility can be excluded, obviously). The varieties which show intensive switching and in which the monolingual MA component diverges from the monolingual speech variety will be later developments, exemplified by the CS variety of Samir and his siblings. The diachronic change of a CS variety takes place both at the individual and the community level. The similarities between the CS varieties of the four Hamadi siblings, as opposed to the other respondents, reveal a certain degree of conventionalisation, albeit perhaps restricted to such a small speech community as the siblings group. On the other hand, the individual differences between the Hamadi brothers with respect to the degree of grammaticalisation *dar* show that idiolectal varieties may develop independently (see the discussion of the *dar* plus infinitive construction in Chapter 6).

The MSA primarily predicts CS patterns that result from the combination of elements from two languages, that is, the CS speech variety in its initial form. When such a CS variety develops independently from the monolingual varieties, this may lead to a new CS variety which challenges some of the predictions of the MSA and of the insertion paradigm generally. However, even where such altered varieties are concerned, most patterns can still be interpreted within the insertion paradigm, and the MSA conveniently detects the points at which a CS variety diverges from insertional CS.

Note that variation and diachronic change of the monolingual language varieties, whether in connection with language contact or otherwise, are not problematic for the insertion approach to CS. The extent to which linguistic variation is taken into account depends upon the level of detail desired. In this study I have chosen to largely ignore individual variation with respect to monolingual language use, using in

preference the abstract notions ‘Moroccan Arabic’ and ‘Dutch’, with the aim of identifying the characteristics of MA/Dutch CS in general. If desired one could replace these abstract notions with others that more adequately fit the speech varieties under discussion, e.g. ‘the MA of second generation immigrants’ or ‘Moroccan adult learners’ Dutch’ or even the speech varieties of an individual speaker at a given point in time.

10.4 Patterns of codeswitching

Insertional CS comes in certain typical patterns, that is to say, not all logically possible insertion types actually occur. The descriptive paradigm influences the interpretation of CS data and the kind of regularities we notice in them. In this section I want to draw attention to some higher-order regularities which the MSA puts in a different light. First of all there is the overwhelming asymmetry in CS: the patterns of embedded Dutch material in MA matrices differ drastically from the patterns of embedded MA material in Dutch matrices. In particular, there are many more Dutch insertions in MA matrices than vice versa. No one familiar with CS data will be surprised by this outcome, but the MSA puts the asymmetry into another perspective because this approach makes no a priori assumption as to which language will function as the matrix language.

Secondly, the investigation of MA/Dutch reveals that, contrary to expectation, embedded Dutch nouns do not have exactly the same distribution as their ML counterparts. More specifically, it turns out that embedded Dutch nouns are almost never modified by a MA attributive adjective. A similar observation can be made concerning another modification relationship, namely that between adjectives and adverbs on the one hand, and degree adverbs that modify these, on the other. This may be related to a further observation: it appears that the embedded Dutch verbs never take MA lexical complements. These observations in fact concern the conspicuous absence of certain insertion types, and they can be subsumed under the term ‘co-occurrence restrictions’.

Sections 4.1 and 4.2 examine the asymmetry and the co-occurrence restrictions in MA/Dutch CS. In the following chapter these issues will be reconsidered from a more cross-linguistic perspective and some explanatory concepts will be proposed for them.

10.4.1 Asymmetric Roles of Moroccan Arabic and Dutch

Because the MSA defines the matrix language on a strictly grammatical basis, this approach makes no a priori assumptions about which language will serve as the matrix language. The data description according to this framework reveals that in the Nijmegen corpus either language at times assumes the role of matrix language. At the same time it is evident that the insertion patterns are entirely asymmetric for the

two matrix languages. Quantitatively the asymmetry is reflected in the fact that the large majority of all insertions consists of Dutch elements in MA matrices. Qualitatively the attested types of Dutch and MA insertions also differ widely. There is a partial overlap between the insertion types in MA and Dutch matrices: in both cases we find embedded subordinate clauses and the insertion of highly specific vocabulary. In the case of the divergent patterns, Dutch insertions are typically content words and have fairly specific meanings, whereas about half of the MA content word insertions are non-specific words like “room” or “woman” which result from a repetition mechanism. In addition, a relatively large number of the MA insertions in Dutch are constituents. Many, but not all, of the embedded MA constituents are Topics, often consisting of independent pronouns. Another recurrent feature of embedded MA NP and PP constituents is that they themselves contain an embedded Dutch noun. Thus many of the MA insertions in Dutch clauses are not at all motivated by the need for specific vocabulary. Moreover, more often than Dutch insertions, MA ones tend to contain function morphemes.

The asymmetric insertion patterns in MA/Dutch are similar to those in other immigrants’ CS varieties. Therefore it is evident that the explanation must be sought in the unequal sociolinguistic status MA and Dutch hold for the respondents. This will be discussed in the next chapter where I will introduce the Community Language/Superimposed Language dichotomy in order to talk about the differential status of the bilinguals’ languages in more general terms.

10.4.2 Co-occurrence Restrictions

Co-occurrence restrictions were found in CS with MA as the ML with respect to adjectival and adverbial modifiers and their heads, and between verbs and their lexical complements. More precisely, the following restrictions on possible CS patterns were noted in the preceding chapters:

- d) Embedded Dutch attributive adjectives very seldom modify MA (ML) nouns; conversely,
- e) MA attributive adjectives rarely modify embedded Dutch nouns;
- c) MA adverbs rarely modify embedded Dutch adverbs or (predicative) adjectives;
- d) Embedded Dutch adverbs never modify MA adverbs or adjectives. (Yet embedded Dutch manner adverbs were found to modify MA verbs.)
- e) In addition, embedded Dutch verbs hardly ever have a MA (ML) lexical complement.

The fact that attributive adjectives are themselves seldom inserted has been noticed several times since Lehtinen (1966). The rarity of EL attributive adjectives is often contrasted with the rather common EL predicative adjectives (Lehtinen, 1966: 227, on American Finnish/English; Bautista, 1980: 36 on Tagalog/English; more examples quoted in Chapter 1, p. 17). To my knowledge, the fact that embedded nouns tend not to be modified by an ML adjective has not been discussed previously within the

framework of an insertion approach to CS.¹⁴ It may be that the latter restriction is peculiar to MA/Dutch, but it is also possible that the matter has not been investigated for other corpora of CS data. If this is the case, it would be a good illustration of how the descriptive method of the MSA brings to light new facts about CS.

The restriction on the insertion of attributive adjectives and on the adjectival modification of embedded nouns can be generalised as a restriction on mixed modifier-head relationships. Similar observations can be made in MA/Dutch for the modification relationship between degree adverbs and the adverbs and adjectives they modify (the above restrictions c and d). I will refer to these restrictions in Chapter 11 as the MODIFICATION RESTRICTIONS. The observation that embedded Dutch verbs do not take MA lexical complements will be referred to as the COMPLEMENTATION RESTRICTION. Note that the analogy of the Modification and the Complementation restrictions is only partial, however: while Moroccan Arabic ML attributive adjectives do not modify EL nouns, Moroccan Arabic ML verbs do occur with Dutch lexical Direct Object and other complements.

The term ‘restriction’ is of course akin to the classical concept of constraints in codeswitching studies. However, unlike the classical CS constraints, the restrictions refer to tendencies rather than absolute rules. Apart from the much attested dearth of embedded attributive adjectives, the restrictions have been noted so far only for this language pair. Hence I do not claim any universality for them, although there are indications that these observations apply to other corpora of CS data as well. In Turkish/Dutch, for instance, the Dutch infinitives that are embedded in the Turkish periphrastic *do*-construction tend not to take Turkish lexical complements either (Backus, personal communication). In the following chapter I will address the relationship between co-occurrence restrictions and collocations, suggesting that both phenomena originate from a common cause.

¹⁴ Within the linear approach, various scholars discussed constraints concerning attributive adjectives, often based on grammaticality judgements, see Chapter 1, p. 17, and Abbassi (1977: 163) on Moroccan Arabic/French.

Chapter 11

Explaining Patterns of Insertion

Eventually the aim of CS research is to explain why CS assumes the form it does. The concept of insertion, to the extent that it is confirmed by the data, provides the first level of explanation. Many of the regularities in CS behaviour can be interpreted and understood through this paradigm and the underlying notion of modularity. Now that the main facts about the morphology and syntax of MA/Dutch CS have been described using the principles of interpretation and classification of the MSA, we are ready to move the investigation to a new level. The MSA identifies a series of regularities in CS behaviour, most of them confirming the notion of insertion; some of them challenging it. Each attested regularity raises new questions and calls for explanation. On the one hand, we need to examine what causes the deviations from the insertion paradigm. What, for instance, explains the occurrence of bare forms, and what determines the word order of embedded attributive adjectives? On the other hand, why do certain insertion types occur while others do not? After all, of all logically possible insertion types, only some are actually attested. Further questions arise when we compare CS in various data sets: why are there systematic differences between data sets from various language pairs, and even between different data corpora of the same language pair? This level of explanation falls outside the competence of the MSA.

Explanations at this level can be approached from two different angles. On one side, CS patterns can be related to sociolinguistic variables in order to account for patterns which are characteristic for entire bilingual speech communities or even types of contact situations. From another angle, the matter can be viewed from the perspective of the individual speaker and explanations can be sought in theories on speech production and on the organisation of the (bilingual) mental lexicon. These two angles may be termed the sociolinguistic and the psycholinguistic explanatory framework. Explanations that relate regularities in CS patterns to language-specific factors would fall under the psycholinguistic type. Much of the CS literature is indeed concerned with sociolinguistic and psycholinguistic explanations. My own interpretation of a number of CS phenomena within the framework of Levelt's (1989) psycholinguistic model of speech production has been published in a number of articles (Boumans, 1995b; 1996; forthcoming). These articles deal, among other things, with the concept of congruence and the insertion of complex word forms.

In this final chapter I will refrain from giving an overview of the explanatory principles that have been proposed in the literature on codeswitching. Rather, I will illustrate how the insertion approach to CS can be applied in both the sociolinguistic and the psycholinguistic domain. I have selected some topics on which the principles

of the Monolingual Structure Approach throw new light, and I will propose possible explanations for some of these. The discussion of these rather heterogeneous topics is explorative in nature: the discussion is also meant to offer outlooks to further research that will examine the validity of the correlations and explanations advanced.

The first two sections deal with the sociolinguistic correlates of CS. Section 1 addresses the asymmetry in CS and the typically asymmetric insertion pattern as identified by the principles of the MSA. In connection with this asymmetry, we will take a closer look at the Moroccan Arabic insertions in Dutch matrices, and more generally, at insertions from the Community Language into Superimposed Language matrices, in section 2.

Following this, we will proceed to discuss some mechanisms in the speech production process that may be responsible for certain CS patterns. The co-occurrence restrictions discussed in the preceding chapter will be reconsidered. I will propose an account for these observations that relates them to the insertion of content word collocations (section 3). Finally, repetition will be discussed as an independent mechanism that can account for some of the less frequent insertion patterns (section 4). Section 5 summarises this chapter.

11.1 Asymmetry in codeswitching

Asymmetry is a striking feature of many CS varieties. Chapters 5-11 show in great detail how the insertion of Dutch elements in Moroccan Arabic matrices differs both quantitatively and qualitatively from the Moroccan Arabic insertions in Dutch. A comparative study of Moroccan Arabic/Dutch in the Netherlands and Algerian Arabic/French spoken in Algeria shows that much of the same asymmetry is found in both CS varieties (Boumans & Caubet, *forthc.*). Moroccan Arabic/Dutch CS largely consists of the insertion of Dutch content words in Moroccan Arabic matrices, whereas the insertion of French content words constitutes the bulk of Algerian Arabic/French CS. In other words, Arabic is the matrix language in the large majority of the CS instances in either text corpus. Both French and Dutch insertions are typically content words with rather specific meanings.

In both CS varieties, the much less frequently embedded Arabic elements share a number of characteristics. Unlike French and Dutch insertions, Arabic insertions are not usually motivated by the need for specific vocabulary items: many of the Arabic insertions are indefinite pronouns or they themselves contain embedded French or Dutch content words (layered insertion). In addition, embedded Arabic elements tend to occur in a clause-initial position, either as an embedded Topic inside the Dutch or French finite clause, or as a clause-external, left-dislocated constituent. Also in both data corpora we find embedded Arabic subordinate clauses, and discourse-emphatic pronouns and some Arabic dis-course markers may accompany Dutch and French finite clauses.

How do we account for the asymmetry, on the one hand, and the fact that Arabic has the same role in both CS varieties, on the other? Unless we want to believe that there is something specific about Arabic that makes it function as the matrix language, the explanation must be sought in the sociolinguistic situation. It is the sociolinguistic situation that accounts for the asymmetry in CS. The comparison of Moroccan Arabic/Dutch and Algerian Arabic/French to codeswitching in closely parallel sociolinguistic circumstances with other language pairs illustrates this. The frequent insertion of Dutch content words is also found in the CS varieties of other immigrant communities in the Netherlands such as the Turks (Backus, 1992; 1996b). Close parallels of Algerian Arabic/French are found in other former French colonies, see for instance De Rooij's (1996) description of Swahili/French CS in the Copperbelt region of Congo-Kinshasa, former Zaire. Consequently, the fact that Arabic is the common matrix language in both Moroccan Arabic/Dutch and Algerian Arabic/French must be due to a comparable sociolinguistic situation despite the quite different status of Arabic in Algeria and in the Netherlands.

In both bilingual communities Arabic is the language 'originally' spoken whereas Dutch and French are acquired in order to communicate with another economically and/or culturally dominant speech community. The dominant positions of Dutch in the Netherlands and French in Algeria are obviously distinct from each other in many ways, but in both cases these languages are needed to gain access to valued information and upward social mobility. Crucially, in contacts between Moroccans and the Dutch in the Netherlands, Dutch provides the medium of communication, and likewise only French is used in Algerian-French contacts. The difference is that the former situation involves much face-to-face interaction while Algerians in Algeria are confronted with French primarily through schooling and (both French and local) media. The asymmetric bilingual situations in Algeria and in the Netherlands can be generalised in a dichotomy based on social distinctions, viz. between the in-group or Community Language and the culturally dominant or Superimposed Language.

11.1.1 Community Language versus Superimposed Language

For a clarification of these terms, let us depart from the postulate that CS takes place in a bilingual speech community; analogous reasonings can be set up for situations of partial bilingualism, trilingualism et cetera. As Romaine notes in the introduction to the 1995 edition of her textbook on bilingualism, "where more than one language exists in a community, they are rarely equal in status" (1995: xiv). Bilingualism on a community-wide scale typically results from the situation that speakers from one speech community massively acquire the language of a culturally and/or economically dominant population, while the reverse does not happen. We therefore distinguish between a language that is used only within the bilingual community, which we call the COMMUNITY LANGUAGE (CL) and a language that is common to both the bilingual

group and the culturally dominant population, to be called the SUPERIMPOSED¹ LANGUAGE (SL). Note that the CL need not be the mother tongue; the Moroccan Arabic/Dutch and Algerian Arabic/French CS varieties, for instance, are spoken by both Arabophones and Berberophones. The SL may be defined firstly as the language which is (predominantly) used in linguistic interactions between members of both speech communities. Furthermore, knowledge of the SL is important for upward social mobility and access to valued information. In the early stages of language contact, the bilingual speakers need the SL only in order to communicate with the culturally dominant population. When the dominant population do not speak the CL, codeswitching plays no role in these contacts. Codeswitching comes into play as soon as bilinguals start to use the SL for communication among themselves.

The speakers of the Community Language may constitute a (local) numerical minority like the Moroccan immigrants in the Netherlands, or they may constitute the numerical majority like the speakers of Arabic in Algeria. The inhabitants of linguistic border areas may at first seem to be examples of a more balanced kind of bilingualism, but here too, prestige factors often lead to the identification of one language as the SL and the other as the CL. In Alsatian/French CS in Strasbourg, for instance (cf. Gardner-Chloros 1991), French is the SL, as the local population is culturally and economically oriented toward France, and Francophone settlers in the Alsace will not bother to speak the local German dialect.

Recall that the CL versus SL dichotomy is logically independent from the ML versus EL dichotomy: the Community Language can be the ML with the Superimposed Language being the EL, and vice versa. However, the sociolinguistic dichotomy makes it possible to predict a number of typical insertion patterns: CL/SL codeswitching is characterised by the insertion of SL content words in CL constituents and, less often, the insertion of certain SL constituents in CL finite clauses. The embedded content words from the SL typically have more specific meanings. Annamalai (1971: 21), for instance, reports on Tamil/English codeswitching that some very frequent English content words like *know*, *go*, *buy*, and *good*, are not inserted. This is formalised by Backus (1996b) in his Specificity Continuum, to which I will return presently. As the investigation of the Moroccan Arabic/Dutch data and their comparison to Algerian Arabic/French in Algeria (Boumans & Caubet, *forthc.*) reveals, the insertion of CL material in SL matrices is less common. Such insertions are less often motivated by the need for specific vocabulary items. CL insertions tend to be more ‘grammatical’ in nature, that is, they are often constituents rather than content words, and they often include CL function morphemes and SL content morphemes. CL insertions have a tendency to occur in clause-initial position of the SL finite clause. The insertion of subordinate clauses and the use of discourse markers from either language seems to be more symmetrical, however.

¹ I avoid the perhaps more suitable term ‘dominant language’ as the latter is traditionally associated with superior linguistic competence.

11.1.2 CS patterns and types of bilingual communities

At first glance, the above predictions seem to hold for most codeswitching varieties, but their validity for CS in various bilingual settings needs to be examined. The social inequality of languages will be more pronounced in some bilingual situations than in others. When the languages have a more equal status, this will probably result in more symmetric CS patterns. A systematic inventory of insertion types will make it possible to relate structural properties of CS varieties to sociolinguistic variables. The establishment of such relationships will be of value for the understanding of linguistic influence generally. The study of the linguistic consequences of language contact in such well-documented social settings like the Moroccan migration to the Netherlands will serve to shed light on the history of languages and peoples in cases where direct historical evidence is lacking.

As a demonstration of this kind of research, some characteristics of CS in three types of bilingual communities will be discussed: immigrant communities in modern industrial societies, colonised societies, and young children growing up in bilingual homes.

Periphrastic do-constructions

Comparing MA/Dutch CS in the Netherlands with Arabic/French CS in Morocco and elsewhere in North Africa brings out one striking difference concerning the way French and Dutch verbs are integrated: embedded French verbs are inflected by means of attaching Arabic prefixes and suffixes to the French verb stem, whereas Dutch verbs, if they occur at all, are embedded in a periphrastic *do*-construction. An explanation for this difference may be sought in language-specific properties of Dutch and French verbs such as their phonological shape or differences in verbal inflection systems. A language-specific explanation seems unlikely, though, if we consider that in the Western Arabic dialects, Arabic inflectional affixes are attached to foreign verbs from various languages, including Spanish in Morocco and Sicilian, Italian and, more recently, English on Malta.² It is plausible that the different sociolinguistic settings in the Netherlands and in North Africa lead to different ways of embedding foreign verbs. As a hypothesis I suggest that the *do*-construction in CS is characteristic for migrant bilingualism in modern industrialised societies. (The bilingual communities which result from present day work migration, such as the Moroccans and Turks in Western Europe, should probably be distinguished from rural settlements like the American Norwegian and American Swedish communities studied by Haugen and Hasselmo, respectively. In the latter situation the immigrants are rather isolated and contact with the Superimposed Language is less intense, at least in the primary stages.)

What links immigrant bilingualism to the periphrastic *do*-construction is the relationship between the intensity of contact and morphological and phonological integration. The degree of morpho-phonological integration is on average inversely

² See the references cited on p. 223, n.2.

proportionate to the intensity of language contact (Ch. 1, section 4.3). In the case of immigrants in modern urban societies contact with the new Superimposed Language is typically sudden and intense. Furthermore it can be argued that the incorporation of foreign verbs by means of a periphrastic construction involves less morphological (and probably phonological) integration than the attachment of ML affixes to foreign verb stems or non-finite verb forms. Thus, periphrastic *do*-constructions may be related to immigrant bilingualism because these constructions involve less morpho-phonological integration and because contact with the Superimposed Language is relatively intense in this situation.

Recall, however, that Moroccan Arabic/Dutch shows no direct correlation between the use of the periphrastic *do*-construction for the insertion of Dutch verbs and the speaker's competence in Dutch (Ch. 9, section 4). It is possible that more psychological aspects of bilingualism also play a role: in the immigrant situation even the speakers who are not particularly fluent in the Superimposed Language will often have the idea that they are supposed to be competent in this language and/or the expectation that they will learn this language in the future. In addition, they will more often assume that their interlocutor is competent in the Superimposed Language.

If there turns out to be a correlation between the way foreign verbs are incorporated and sociolinguistic factors, any explanation for a given insertion routine should take this into account. Clearly such a correlation would be no more than a tendency since other factors are also relevant.

The correlation between ways of incorporating foreign verbs and sociolinguistic factors is obscured by the fact that various (Community) languages display only one routine, irrespective of the social circumstances. Periphrastic *do*-constructions, for instance, are a pervasive feature of the monolingual varieties of many Turkic and Indo-Iranian languages, among others. In CS with Turkish as the Community Language, for instance, we will see that verbs from the Superimposed Language are embedded in a *do*-construction irrespective of the social setting, i.e., both in migrant communities (cf. Backus, 1992, 1996b, on Turkish/Dutch) and in Turkish speaking lands (cf. Rudin & Eminov, 1990, on Turkish/Bulgarian). For the relationship between sociolinguistic variables and ways of embedding foreign verbs to become apparent, the same Community Language must obviously show various ways of verb-embedding in different language contact situations. (Recall that the insertion of CL content words in SL matrices is typically infrequent.)

Asymmetry in the insertion of content words

Backus (1996b), in his study of Turkish/Dutch in the Netherlands, argues that code-switching in the early stages of language contact consists in the insertion of highly specific lexical items. He places all lexical items on a gradual scale of specificity³

³ Backus' notion of specificity includes both semantic specificity ('entrenchment potential') and structural autonomy. With respect to the diachronic stages of content word insertion, semantic specificity seems to be more central, however. It

which he calls the Specificity Continuum (1996b: 115-31). He presents the following diachronic view on CS varieties: in the earliest stages of, e.g. immigrant bilingualism, only the most specific lexical units are inserted from the new language, notably proper names. Subsequently, with the generations who make greater use of this new language, speakers will also insert less specific units, such as words for concepts that are not lexicalised in the ML. These words are called cultural borrowings filling lexical gaps, in Myers-Scotton terminology (1993b). But eventually, nearly any EL content word can be considered as filling a lexical gap because, Backus argues, “translation equivalents are often not exact equivalents, because the encyclopaedic knowledge that a speaker has about a certain word will seldom exactly match the connotations which the ML hold for her” (1996b: 120).

The insertion of highly specific lexical items is, however, very asymmetric in occurrence. In Turkish/Dutch as well as MA/Dutch the large majority of all CS instances consists in the use of Dutch content words, particularly nouns. The opposite is very rare: ten hours of recordings yielded only a handful of embedded MA content words in Dutch constituents. Six of these indeed refer to highly specific Moroccan or Islamic cultural concepts; the other ones result from a repetition mechanism (on which see section 4 below).

What might account for this asymmetry is the fact that the Turkish and Moroccan immigrants are used to speaking Dutch primarily with people who do not know any Turkish or Arabic. This might possibly block the emergence of a variety of Dutch that includes Arabic or Turkish words, even for purposes of in-group communication. However, the asymmetry is not much different in a situation where everyone speaks the Community Language, such as in Arabic/French CS spoken in Morocco and elsewhere in North Africa. Even though the Arabic/French bilinguals in Morocco have relatively little occasion to speak to monolingual Francophones, it is just as uncommon to insert Arabic content words in French matrices in MA/French CS in Morocco. As a rule, so it seems, content words from the Community Language are not embedded in Superordinate Language constituents and clauses. It may be argued that the relatively low prestige of the Community Language blocks insertions from that language. This can be compared with the typically low number of borrowed lexical items in the case of substratum interference (community scale shift to a prestigious language variety) as opposed to superstratum interference (shift to a less prestigious variety), in Thomason & Kaufman’s (1988: 116) terminology. However, the relevance of prestige is not so obvious in view of the other forms of Community Language/Superimposed Language CS that do occur in the same conversation.

Kulick & Stroud (1990) present an interesting case in point. They study code switching in Gapun, a tiny isolated village in the Sepik region of Papua New Guinea. The local population is rapidly shifting from the local language Taiap to Tok Pisin,

must be semantic specificity, rather than autonomy, which accounts for the observation by some researchers that very frequent lexical items are less likely to occur as EL forms than other content words of the same class (cf. Annamalai, 1971: 21).

the country's main lingua franca (cf. Kulick, 1992). While the younger population has no active mastery of the local vernacular, Taiap/Tok Pisin CS is a common feature of the adults' speech. One of the forms Taiap/Tok Pisin CS takes, is the insertion of Taiap content words in Tok Pisin utterances: "certain words, especially nouns denoting everyday items such as betel nut, sago, fire, basket, coconut, water and so on, are likely to be named in the vernacular, even if the rest of the utterance is in Tok Pisin" (Kulick & Stroud, 1990: 212). The relative status of Taiap and Tok Pisin in Gapun village appears to be comparable to that of Arabic and French in Morocco, although no language shift is involved in the latter case. This raises the question of why content words from the Community Language are inserted in Superimposed Language utterances in Taiap/Tok Pisin CS, but not in Moroccan Arabic/French. Note that also in Morocco French and MA/French CS are used in casual, everyday conversations in certain milieus (cf. Slaoui's 1986 corpus of table conversations). It can be hypothesised that different attitudes toward bilingualism are at stake in Morocco and in the Papuan village of Gapun; however, this calls for further investigation.

Symmetrical CS in bilingual households

Several scholars describe CS in the speech of young children (often their own) who grow up in bilingual households (e.g. Petersen, 1988; Kaufman & Aronoff, 1989; Choi, 1991; Kwan-Terry, 1992). At home the language of (one of) their parents is spoken alongside the language of the society outside the home, which happens to be English in the cited references. While there will be significant differences among the described households, the sociolinguistic situation of these children appears to diverge from the 'classical' immigrant communities in two ways: firstly, the children are addressed by (one of) their parents in English, and secondly, the bilingual community of which they are members is restricted to the family. Choi (1991: 879), for instance, reports that his two daughters communicate exclusively in English with their New York peers, so that Korean/English CS is restricted to use within the home.

The CS varieties of these children do not reflect the 'normal' Community Language/Superimposed Language dichotomy. Rather, their CS seems to be more symmetric with the insertion of content words in either language. Consider two examples produced by Choi's daughters: both the Korean and the English word for "hand" are inserted.

- (1) *I don't eat with a son*
hand
 "I don't eat with a hand." Korean/English (Choi, 1991: 889)
- (2) *ikē kunyang wuli hand-tulwu hae-ss-ē*
this just we hand-with do-PAST-DECLARATIVE
 "We just did this with hands." Korean/English (Choi, 1991: 892)

(3) ngoh yiu nei *put back* goh *knife*, wui *cut* nei ga
 “I want you to put back that knife; it will cut you!”
 Cantonese/English (Kwan-Terry, 1992: 249)

The fact that some of the ‘home varieties’ of CS are more symmetrical than, for instance, the CS varieties described for the Turkish and Moroccan migrant communities in the Netherlands is probably due to the relatively equal status of both of these children’s home languages, at least at the time of the recordings. It should be noted that the sociolinguistic environment of bilingual children is often subject to rapid change. Some of the bilingual household children whose language behaviour has been studied seem to shift completely from the language of their parent(s) to the language of their peers (i.e. English) as soon as they begin to develop outdoor activities. Such a rapid shift may be related to some of the atypical CS patterns described by Petersen (1988) and Kaufman & Aronoff (1989). In particular, preceding a complete shift to English (the SL of the outdoor community), the children they study speak English using some content words from the home language (the CL). Myers-Scotton (1997: 226) does indeed speak of a change in the ML for the case described by Kaufman & Aronoff, as well as for a similar case described in Kuhberg (1992). Choi and Kwan-Terry do not report any language shift in the case of their children, however.⁴

Most attention in CS studies has been devoted to the insertion of Superimposed Language material in Community Language utterances, which constitutes the majority of CS data. This is not to say that insertions from the Community Language have not been considered at all in the literature on CS, but the MSA views this from a different

⁴ Treffers-Daller (1995) discusses another complex sociolinguistic situation which yields atypical and more symmetric CS patterns: the case of adolescents of Turkish descent who had grown up bilingually in Germany and went to Turkey for the purpose of education.

angle. In particular, the fact that the MSA defines the matrix language on a very local basis, i.e., the finite clause and the constituent, leads to the identification of very local changes of the matrix language in a text. In this section I want to draw attention to the regularities which the MSA classifies as insertions from the Community Language.

11.2.1 The Matrix Language Turnover

The description of MA/Dutch CS shows the insertion of Dutch content words and constituents in MA matrices, and relatively small numbers of MA insertions in Dutch matrix constituent and matrix clauses. Some respondents produce insertion patterns of both types, but an implicational hierarchy exists regarding with respect to the amount of Dutch used in mixed utterances: only those respondents who insert Dutch content words, may also insert Dutch constituents; only those respondents who insert Dutch constituents may also insert MA content words and constituents in Dutch clauses. The size and complexity of Dutch elements used in mixed utterances is roughly proportionate to the amount of Dutch the respondents use in their daily communicative interactions (and, by implication, their competence in this language), so beginning learners like Mustafa are only able to insert some Dutch content words, while advanced learners like Hayat and respondents who have grown up in the Netherlands occasionally insert larger chunks of Dutch, and even produce Dutch clauses containing MA insertions.

With respondents who make intensive use of Dutch in their everyday life, embedded Dutch content word morphemes occasionally go together with other Dutch morphemes and manifestations of Dutch grammatical rules. This is reflected in CS by the insertion of Dutch content word collocations and constituents. When Dutch verbs are inserted, as is common in some speakers' CS varieties, this can have considerable syntactic consequences within the MA matrix clause, as I discussed at the beginning of Chapter 10. If the selection of a Dutch verb goes together with Dutch verbal inflection, then the clause as a whole will be considered as a Dutch clause and will also have Dutch constituent order, as the MSA predicts. In this case, however, we may still find manifestations of MA grammar within the Dutch clause, namely in the form of MA constituents. These MA constituents are syntactically embedded because they occupy a position in the Dutch clause. At this point we can say that the ML has changed from MA, the 'default' ML in MA/Dutch CS generally, to Dutch in a number of instances. Using a term coined by Myers-Scotton (1993b) I refer to this change in matrix language as the Matrix Language Turnover.

This inversion of the ML/EL relationship can be generalised to CS in other bilingual communities. Assuming that the bilingual community speaks the Community Language first before it comes in touch with the Superimposed Language, the earliest forms of CS consist of inserting (highly specific) content words from the latter into matrices from the former language. Hence the Community Language is the first ML. The Matrix Language Turnover takes place when the bilingual community members

(5) *the customer fills forms and surrenders* kiasi fulani ch-a pesa say like
 amount some CL7-of money
200 shillings every month for two years
“The customer fills forms and surrenders some amount of money say like 200
shillings every month for two years.”
Swahili/English (Myers-Scotton, 1993b: 72)

11.2.2 Community Language insertions

With respect to the MA/Dutch text corpus, the ML turnover is instantiated by three types of MA insertions in Dutch matrices: a) the MA insertions which result from the repetition of content words or constituents; b) the insertion of highly specific

⁵ Myers-Scotton (1993b: 70-4) acknowledges both synchronic and diachronic shifts in the ML, but because the ML is defined differently in her MLF model, her Matrix Language Turnover does not exactly match the concept advanced here. Example (5) above, for instance, does not exemplify a turnover of the ML in Myers-Scotton's view. "According to the MLF model, [Swahili is the ML in this clause, and] *the customer fills forms and surrenders* is an EL island, as are *say like 200 shillings, every month, and for two years*" (1993b: 72).

Islamic and Moroccan vocabulary items (Ch. 9, section 1.1); and c) ‘other’ embedded MA constituents. These insertion types reflect three different codeswitching mechanisms. In section 4 below we will discuss repetition as a distinct CS mechanism. The culturally specific MA lexical items which occur in Dutch clauses contrast with the embedded MA constituents in a number of ways: Firstly, the former have very specific lexical meanings, while this does not characterise the EL constituents. Secondly, the culturally specific items are usually nouns, although they sometimes trigger the insertion of a full NP; the MA embedded constituents, on the other hand, are nominal, prepositional or adverbial constituents and quite often they are pronominal in nature. Thirdly, the culturally specific items tend to occur in the focused position in the Dutch clause, whereas the majority of the EL constituents are found in Topic position. Therefore, the insertion of culturally specific items parallels the insertion of highly specific Dutch content words in the first stages of MA/Dutch bilingualism; compare Backus’ view on the diachronic development of CS varieties sketched in section 1.2 above. This type of MA insertion represents the first (and possibly also the final) stage of CS behaviour which starts at ‘the other side’, that is, by inserting MA elements into Dutch clauses. The insertion of typically non-specific MA constituents, on the other hand, can be viewed as the last manifestations of what was once the matrix language, after increasingly large chunks of Dutch are inserted, up to the point where Dutch must be recognised as the matrix language of the finite clause.

The last manifestations of Community Language grammar

If we regard embedded MA constituents as the last foothold of the former matrix language grammar after the ML Turnover has taken place, we can understand why these EL constituents are so different from the EL constituents from Dutch. Myers-Scotton (1997: 223) writes: “the ML is the language projecting the morphosyntactic frame for the entire CP which shows intrasentential CS”. In most instances of MA/Dutch intrasentential CS, MA can indeed be said to project the morphosyntactic frame for the entire CP, or at least for the finite clause. When MA constituents are inserted into Dutch finite clauses, MA no longer projects the morphosyntactic frame on the finite clause level and there is a local turnover of the matrix language. However, MA still projects the morphosyntactic frame of this single constituent. I view this as the continuation of MA’s role as the ML on a very local level, namely that of the nominal, prepositional or adverbial constituent.

Let us reiterate the differences between the Dutch and MA constituents once more: Dutch EL constituents can be divided into two major types: prepositional and adverbial constituents in peripheral syntactic (adjunct) positions, and constituents built around Dutch content word heads in focused positions, notably complements of verbs and predicates in copular constructions. Dutch EL constituents are typically rich in content and high on the specificity continuum. Embedded MA constituents, on the other hand, often - although not exclusively - occur in topicalised positions and are not necessarily specific or rich in content. As Topics, embedded MA

constituents typically refer to given information and consequently they are often realised as pronominal forms. A final difference between Dutch and MA embedded constituents is that Dutch function morphemes such as determiners tend to be avoided (or omitted) in Dutch constituents which occupy argument positions (the Dutch constituents in adjunct positions are often fixed expressions, e.g. *zo nu en dan* “every now and then”). MA embedded constituents, on the other hand, often consist entirely of MA function morphemes or of MA function morphemes and Dutch content words. Reconsider some of the examples cited in Chapter 9, reproduced here. In the first two examples, the MA constituent occupies a Topic position: in (6) the Topic is clause-internal; because it does not assume the Subject function itself, the place adverb *hna* “here” triggers the inversion of Subject and finite verb in the Dutch clause. In (7), MA *hadak š-ši* is a clause-external, left-dislocated Topic. This is marked grammatically by the Dutch resumptive pronoun *dat*. (8) exemplifies another typical pattern of MA insertions in Dutch: layered insertion. Here we see a MA possessive PP, the lexical core of which (i.e., the possessor) is realised by a Dutch content word. Clearly, these MA insertions are not motivated by semantic specificity of the lexical items. Note that these MA constituents, although atypical insertion types when the Community Language is the ML, are very transparently embedded in the syntactic sense of the word: they form part of Dutch grammatical structures.

- (6) *iyeh, hna gaan hun met de tijd mee*
 yes here keep·pace· they with the time ·keep·pace
 “Yes, here they keep pace with time.” (Hayat, ex. (42) in Ch. 9)
- (7) *hadak š-ši, dat is wat anders*
 DEM DEF-thing that is something different
 “As for this, that’s something different.” (Abdellah, ex. (31) in Ch. 9)
- (8) *ik ben niet tevreden over eh ____ kwaliteit dyal eh, ja, dyal ____ faculteit*
 I am not satisfied about er [the] quality of er well of [the] faculty
 “I’m not satisfied with er the quality of er, well, of the faculty.”
 (Hocine, ex. (41) in Ch. 9)

Note that MA embedded constituents, unlike Dutch ones, are often indefinite pronouns, as illustrated by *hadak š-ši* “this” in (7) and *ši waḥed* “someone” in (9) below.⁶

⁶ These EL constituents should not be confused with the personal pronouns that mark Topic shift in Arabic (cf. Ch. 3). The latter are extra-clausal and never occur as the first constituent within a Dutch finite clause. Since Arabic is a ‘pro-drop’ language, syntactically embedded Arabic pronouns are usually indefinite rather than personal pronouns.

- (9) ana ma ʕemmr-i ši wahed heeft mij gediscrimineerd
 1SG NEG ever-1SG some one has me discriminated
 “As for me, nobody ever discriminated against me.”
 (Hayat, ex. (29) in Ch. 9)

The observations with regard to MA/Dutch can be generalised to other codeswitching varieties by means of the Community Language/Superimposed dichotomy. The features of the MA EL constituents are characteristic of embedded constituents from the Community Language in general. Some examples will demonstrate that the MA/Dutch CS patterns can be translated into Community Language/Superimposed Language patterns. First consider four examples from Moroccan and Algerian Arabic in contact with French and Spanish, which yield quite close parallels to MA/Dutch.

- (10) *c'est, c'est un signe de modernité que de parler .. que* l-wahed parle,
 it's it's a sign of modernity that of speaking that DEF-one speaks
parle le français
 speaks DEF French
 “It's, it's a sign of modernity to speak, that one speaks, speaks French.”
 Moroccan Arabic/French (Rhaïb, 1996: 158)

- (11) had š-ši no merece la pena
 DEM DEF-thing not be-worth the trouble
 “This isn't worth the trouble.”
 Moroccan Arabic/Spanish (Herrero Muñoz-Cobo, 1996: 146)

- (12) kaš nhar, wahed il s'est plaint, ula?
 INDEF day someone he him-REFL-is complained or
 “Has anyone ever complained, or what?”
 Algerian Arabic/French (Boumans & Caubet, forthc.)

- (13) ɾa-h ɕa reflète la personnalité dyal quelqu'un
 PRES-3M this reflects the personality of someone
 “This reflects someone's personality.”
 Moroccan Arabic/French (Wernitz, 1993: 340; ex. (41) in Ch. 1)

The first two examples are very similar to (9) above: the clause starts with an indefinite pronoun “someone” or “this” in Arabic, and continues in the Superimposed Language (Dutch, respectively French or Spanish); the Arabic indefinite pronoun is embedded in the SL finite clause where it takes on the Syntactic function of Subject. Example (12) resembles (9) in that both begin with an Arabic temporal adjunct, but in (12), the ensuing Arabic indefinite pronoun remains external to the French finite clause; it is realised as a left-dislocated NP, just like *hadak š-ši* “this” in the MA/Dutch example (7). More examples of embedded MA pronouns are cited in Chapter

2 (exs. (48) and (49)). The MA/French example (13) is analogous to (8) above. This type of layered insertion is characteristic for CS with the Superimposed Language as the ML: the embedded Community Language NP or PP contains a Superimposed Language lexical core. Further MA/French examples of this type have been cited in previous chapters (exs. (25) in Ch. 1, (36)-(38) in Ch. 2). Similar cases are found in other language pairs, which suggests that we are dealing here with a recurrent feature of codeswitching:

- (14) a *druge* *will throw me a party there*
 and others·F
 “And the others will throw me a party there.”
 Serbian/English (Savić, 1995: 483)
- (15) *ei se did’n notice että*
 not he didn’t notice that
 “He didn’t notice that.” Finnish/English (Lehtinen, 1966: 171)
- (16) *and some schools are conducting review classes* *para sa* *kanila*
 for DIRECTIONAL them
 “(..) and some schools are conducting review classes for them.”
 Tagalog/English (Bautista, 1975: 83; 1980: 218; ex. (30) in Ch. 1)
- (17) *aku vind die nieuwsgierigheid die echt te gek*
 I find this curiosity this really too mad
 “I really find this curiosity fantastic!”
 Moluccan Malay/Dutch (Voigt, 1997: 69)
- (18) (*yu wok. yu wok. no ken toktok planti,*) *yu-yi wokim wok tasol.*
 you work you work NEG can talk plenty you-ERG work work just
 “You work. You work. You can’t complain, you just do the work.”
 Taiap/Tok Pisin (Kulick, 1992: 77)

As mentioned, the insertion of Superimposed Language constituents which contain function morphemes is rather constrained, at least in CS with genetically and typologically diverse languages. As possible explanations we might conjecture that lack of congruence blocks certain EL constituents from the Superimposed Language or that the syntactic procedures which assemble such EL constituents are not active in an ML environment. Comparable restrictions do not apply to EL constituents from the Community Language, however. The latter may freely include function morphemes such as paradigmatically organised personal pronouns, determiners and case markers. The above examples (15)-(18) show a Community Language personal pronoun in a Superordinate Language matrix clause; the Serbian form *druge* “others·F” in (14) and the Finnish *että* “that” are paradigmatically organised indefinite pronouns. In (16), an English clause contains the Tagalog PP *para sa kanila* in which

kanila is a case-marked form of the third person plural pronoun. The final example shows Tok Pisin, the lingua franca of Papua New Guinea, as the ML and Superimposed Language; the Community Language is Taiap, a Papuan language spoken in a tiny village in the Sepik region (Kulick, 1992). Note that in the last Tok Pisin clause of (18), the personal pronoun is realised in Taiap and accordingly marked for ergative case (whereas Tok Pisin is a nominative/accusative language).

From the examples cited above the impression emerges that embedded constituents from the Community Language typically (although not exclusively) occur at the beginning of a clause or sentence. Such a trend is firmly established for the MA/Dutch data base. In order to explain this phenomenon, I envisage three approaches. The first is a division of labour between the Community and the Superimposed Language in terms of Given versus New information. The second approach assumes a direct connection between the Community Language and Topics. Finally the matter can be considered from a linear point of view: speakers tend to start in the Community Language and continue their utterance in the Superimposed Language.

Given versus new information

The division of roles between the Community and the Superimposed Language along the lines of Topic/Focus or Given/New was first commented on by Hasselmo with respect to American Swedish. His Ordered Selection model (see Chapter 1) is designed so as to allow for Swedish Subject NPs to be followed by an English VP, but not vice versa: if the Subject is realised in English, the sentence must be completed in English. He notes that CS often takes place “before a direct object or a predicative adjective or noun” (i.e., these elements are often inserted from English, the Superimposed Language). “This pattern,” he continues, “may have to do with the tendency to locate the units that convey new information in these positions” (Hasselmo, 1975: 259). A more basic explanation is offered by Boeschoten & Verhoeven (1987: 209) who point out that embedded nouns “most often designate (concrete) objects and seldom persons”. Inanimate nouns are seldom agents and therefore occur less often as Subjects, at least in languages where subjecthood is related to agentivity.

Both Hasselmo’s distinction between new and given information and Boeschoten & Verhoeven’s explanation are covered in the trend to insert highly specific lexical items from the Superimposed Language, formalised in Backus’ (1996b) Specificity Continuum. Lexical items with specific meanings will often designate inanimate objects, which are less eligible as Subjects, or be proper names, which are not usually counted as instances of codeswitching. Furthermore, specific lexical items will tend to occur as new information in a focused position; given information is often referred to by pronominal forms, as are Topics, since Topics usually represent given information. True, embedded content words are more frequent in focused positions, but this probably holds for matrix language content words too. As yet there is no indication that EL content words occur in focused positions more often than ML content words.

If the specificity of embedded lexical items from the Superimposed Language accounts for the observation that elements from this language tend to occur in focused position, it does not automatically follow that everything except the Topic will tend to be in the Superimposed Language. In other words, specificity does not explain why Topics tend to be in the Community Language even if the rest of the clause is in the Superimposed Language. In order to account for this a more general division of labour in terms of Given/New must be assumed which associates the Topic position with the Community Language.

Topics

Embedded constituents from Moroccan Arabic, and from the Community Language generally, often occur in Topic position. This was demonstrated for MA/Dutch in Chapter 9, and many of the examples reproduced above likewise show Community Language constituents in Topic positions, either clause-internal (often assuming the grammatical function of Subject) or clause-external as left-dislocated constituents. To this we may add the foregrounding strategies discussed in Chapter 3: the foregrounded element is a particular type of Topic. While prototypical Topics are given information and hence usually realised as attenuated pronominal or zero forms, foregrounding mechanisms are used to highlight a contrastive or newly (re-)introduced Topic. Remember that in CS with Arabic, Arabic personal pronouns commonly precede a clause in the Superimposed Language with the function of marking Topic Shift. In Japanese/English, there are many instances of the Japanese Topic-Comment structure in which the Comment is realised in English while the Topic is in Japanese and/or marked as such by the Japanese Topic marker *wa*.

- (19) ?iħna *it is none of our business*

1PL

“It’s none of OUR business.”

Levantine Arabic/English (Barhoum, 1994: 100; ex. (38) in Ch. 3)

- (20) *Powell street wa we used to call it Little Tokyo*

TOP

“As for Powell street, we used to call it Little Tokyo.”

Japanese/English (Nishimura, 1986: 135; ex. (27) in Ch. 3)

Sobolewski (1982), on Tagalog/English, cites two examples that bear a superficial resemblance to the ‘pronoun doubling’ in CS with Arabic: *ako, I miss you a lot!* and *ako, I just wear old dresses ...*, where *ako* is the Tagalog 1SG TOPIC pronoun. According to the author, *ako* adds emphasis, and “in neither case does *ako* have any syntactic relationship with the verb” (1982: 50).⁷ In addition, recall the discussion

⁷ The similarity with Arabic may be merely superficial. I do not know whether these examples reflect any foregrounding mechanism of monolingual Tagalog.

of Bautista's (1975, 1980) Model of Bilingual Competence in Chapter 1 (section 3.3.2). Bautista found that Tagalog constituents which function as an NP in English clauses are always marked as Topics. I will return to Bautista's finding shortly.

Kulick & Stroud (1990: 222) give a quantification of Taiap/Tok Pisin code-switching in their text corpus. It appears that switches "between a clause and an element positioned outside that clause through either topicalisation or left-dislocation" (1990: 223) make up a large part of all intrasentential CS (excluding the insertion of single words).⁸ Although they do not state that in all these cases the topicalised or left-dislocated element is in Taiap, the Community Language, this is true for the examples they cite.

- (21) *hgrag-rɛ, toktok wantaim em*
 evening-LOC talk with 3SG
 "In the evening, talk with him .."
 Taiap/Tok Pisin (Kulick & Stroud, 1990: 224)

- (22) *yu ani les-tɛt-hjan bai yu likim desela samting*
 2SG who tired-2SG-REL FUT 2SG see this thing
 "You who are tired, you are going to see this thing."
 Taiap/Tok Pisin (Kulick & Stroud, 1990: 224)

In a word, we observe a tendency for Topics, whether highlighted or attenuated, to be realised in the Community Language. Conversely, embedded elements from the Superimposed Language (mostly content words rather than constituents) are concentrated in focused positions. With respect to the Topic-Comment construction Japanese/English, for example, Nishimura remarks that "Topics are realized only in Japanese not in English" (1989: 376). It should be noted, however, that this is merely a trend, and there are many counter-examples. Poplack's (1980) study of Puerto Rican Spanish/English CS in New York, for instance, shows only a relative preference for Community Language Subject constituents: she counts 44 Spanish Subject NPs in English clauses, as against 25 English ones in Spanish clauses; embedded Object NPs occur slightly more often in English: 78 against 62 Spanish ones (1980: 602).⁹ (I assume the Subject and Topic constituents in these data coincide to a large extent.)

⁸ They count 8 tokens of this, out of a total of 41. The large majority concern switches between a main and a subordinate clause (18 tokens) or switches "between two independent clauses" (1990: 223); the examples suggest that cases of reported speech are meant here. This leaves few instances of other types of intrasentential CS. Singly inserted words are not counted.

⁹ Regarding the interpretation of *Table 2* in Poplack (1980) see the remark on p. 18 in Ch. 1.

As a final example, consider the following extract from a theatre performance by the Algerian comedian Mohamed Fellag, registered on video¹⁰ in Algiers in 1990 (Caubet, 1997). In one of his sketches, Fellag depicts a situation in which the Islamic fundamentalist movement has come to power in Algeria, and the government has bought the Berlin wall from re-united Germany in order to separate the men and women from each other. Following the separation, Fellag narrates, the feminine society thrives due to the women's commitment; the men, on the other hand, are not used to look after themselves and pauperise:

(23) *et comme* *n-nsa*, *elles sont plus organisées* *ʕli-na*, *elles*
and since DEF-women•PL they•F are more organised above-1PL they•F

sont plus courageuses, elles sont plus productives, *y-ʕud y-kun*
are more courageous they•F are more productive 3-return 3-be

ʕend-hūm kūlleš: *l'agriculture* *te-t-mešša* (..)
at-3PL everything DEF-agriculture 3F-MP-go

r-rġal, xlaš, ils en peuvent plus, ra-hūm ġir y-ramp'i-w f
DEF-men finished they•M of-it can no•more PRES-3PL just 3-crawl-PL on

l-arḍ

DEF-ground

“Now the women, since they are more organised than us [men], they are more courageous, they are more productive, they will have everything again: agriculture flourishes, (..) As for the men, it's over, they can't take it anymore. They're just crawling on the ground.”

Algerian Arabic/French (Caubet, 1997)

Algerian Arabic is used to set the topic on which a number of predications are made: *n-nsa* “the women” in first line and *r-rġal* “the men” in the last function as left-dislocated Topics in French matrix sentences. Note that in this passage, the left-dislocated Arabic NPs *n-nsa* and *r-rġal* do not only mark the Topic for the ensuing finite clauses, which happen to be French, but for longer stretches of narrative discourse that relate the conditions of the women and men, respectively. Recall that in Chapter 3, clause-external foregrounding mechanisms like left-dislocation were regarded as discourse markers in the sense of Schiffrin (1987). According to her, “markers bracket units of talk. Sometimes those units are sentences, but sometimes they are propositions, speech acts, tone units” (1987: 35).

¹⁰ Fellag, Mohamed (n.d.) *SOS Labess. Algérie: pour le meilleur et pour le rire*. Alger, Le Petit Théâtre & TÉLÉCINEX.

Linear switching

The proclivity of the Moroccan Arabic insertions to occur in the topical position of Dutch matrix clauses may also be accounted for by a much more down-to-earth explanation: the MA insertions occupy the first position simply because there is a left-to-right switch from MA to Dutch. That is, the speaker starts her utterance in MA and then continues in Dutch. When the already uttered portion of MA has the form of a nominal, prepositional, or adverbial constituent, it may become integrated as the Topic constituent in the ensuing Dutch finite clause. If the clause-initial constituent is an NP, it may, moreover, function as the Subject of the Dutch clause. In addition, a clause-initial NP can be integrated as a left-dislocated constituent, which is grammatically marked in the Dutch clause by means of a resumptive pronoun.

A supporting argument for this explanation is that it is not just MA Topics that precede Dutch finite clauses: we also made the observation that various MA discourse markers tend to occur in clause-initial position, notably the question marker *waš* and causal and adversative conjunctions (Chapter 9, section 3.2). Besides, MA clauses that function as a constituent in a Dutch finite clause tend to occur in clause-initial position too. Recurrent examples of this are MA conditional clauses (Chapter 9 section 3.1.2). Dutch elements, on the other hand, do not often occupy the first position of MA clauses, or precede these as extra-clausal elements, although some Dutch discourse markers and embedded clauses constitute an exception (Chapter 8). Thus the insertion of MA Topic constituents in Dutch finite clauses might be part of an overall tendency in some speakers to start an utterance in MA and continue in Dutch. The switch may or may not be accompanied by an interruption of the speech flow. In this context the following examples of linear switching uttered by Samir are interesting. The Dutch finite clauses in these two examples lack an overt Subject constituent, yet person and number are marked in the MA onsets. (Recall that the MA free form pronouns like *huma* in (24) are markers of Topic shift rather than Subjects). Similar cases of rephrasing in Dutch were cited in Chapter 9, exs. (27), (28) and (48).

- (24) ka-neʃref l-luġa dyaḥ-hūm u huma ka-y- eh zie-n mij als een
 ASP-1-know DEF-language of-3PL and 3PL ASP-3- er regard-PL me as a

vreemde

outsider

“I know their language but yet they regard me as an outsider.” (Samir)

- (25) kifaš ġadin y- controler-en of wat er in de boeken staat,
 how FUT·PL 3- verify-PL whether what PARTICLE in the books is·written

of dat waar is?

whether this true is

“How will they verify whether what’s written in the books, whether this is true?” (Samir)

For some speakers, like Samir in the above examples, it is plausible that they tend to start in MA and continue in Dutch due to limited competence in MA, combined with a strong motivation to speak this language. They start in the Community Language, possibly motivated by a desire to express ethnic solidarity, but have to continue in the Superimposed Language in which they are more fluent because formulating in the Community Language takes too much time and they run the risk of losing the floor if their interlocutors are impatient. Note, however, that in the MA/Dutch data the tendency to switch to Dutch (in a linear sense) after the first word or constituent is not restricted to Dutch-dominant respondents (cf. examples by Hayat reproduced in (6) and (9) above).

Interestingly, the left-to-right processing explanation for the tendency to insert MA Topic constituents in Dutch finite clauses recalls something of the Equivalence Constraint proposed by Poplack in the early 1980s. While I reject the idea that equivalent word order or constituent order is a general organising principle in CS, it may turn out that linear CS plays a role in explaining certain patterns. Note, however, that the MA constituents that occur in the Topic position in Dutch finite clause are still embedded in the syntactic sense of the word: The Dutch finite verb agrees with the MA constituent in person and number if it is an NP that can function as the Subject or, if the MA constituent is not the Subject, its occurrence in the first position leads to the inversion of finite verb and Subject in the Dutch clause (i.e., the ‘verb-second’ rule applies).

To conclude, the Moroccan Arabic/Dutch data suggest that there is an overall tendency to start an utterance in the Community Language and to continue in the Superimposed Language. The question is whether there is also another, independent tendency for Topic constituents to be in the Community Language even if the rest of the clause (or sentence or utterance) is in the Superimposed Language. This is difficult to establish since Topic constituents are often also first constituents. However, Bautista (1975, 1980) provides some evidence that we are indeed dealing with two independent phenomena. She made the observation that Tagalog constituents that occupy the position of an NP in English sentences always take on the so-called ‘*ang*-form’. The ‘*ang*-form’ is the Topic form of Tagalog NPs according to Schachter & Otnes’ (1983) Tagalog grammar, labelled Subject by Bautista, see Chapter 1, section 3.3.2. These Tagalog Topic constituents occur clause-initially as the Subject of an English clause but also clause-finally as complements of English verbs or prepositions (Bautista, 1980: 57-60).

11.3 The Modification and Complementation Restrictions

In the preceding chapter I introduced the terms ‘modification and complementation restrictions’ to refer to the observation that certain patterns of co-occurrence of MA and Dutch lexical items are conspicuously absent from the data. The modification

restrictions refer to the observation that a) EL attributive adjectives very seldom modify ML nouns; b) ML attributive adjectives seldom modify EL nouns; c) ML adverbs seldom modify EL adverbs or (predicative) adjectives; d) EL adverbs never modify ML adverbs or adjectives. The complementation restriction refers to the fact that embedded Dutch verbs never occur with a MA (ML) lexical complement. Interestingly, the formal relationships involved in these ‘co-occurrence restrictions’ between Dutch and MA lexical items are the same as those found in the most common types of embedded collocations: adjective-noun collocations and verb-object collocations. This creates the impression that the many EL collocations somehow exclude the engagement of EL items in modification or complementation relationships with ML lexical items. Possibly the two phenomena have a common explanation. This possibility will be explored in the present section.

Note that the co-occurrence restrictions are observable only when Moroccan Arabic is the ML; they do not seem to apply when Dutch is the ML. I attribute this difference not to language specific differences between MA and Dutch, but rather to their differential social status in the bilingual community. As a result of this social asymmetry, few MA content words are inserted in Dutch matrices. In the instances where MA content words are inserted, repetition turns out to be an important factor which appears to override the co-occurrence restrictions. I will return to this shortly. First I will elaborate on the significance of collocations for a number of CS patterns.

11.3.1 Collocational ties

The Nijmegen data contains many examples of mainly Dutch inserted collocations such as *lagere school* ‘primary school’ and *rekening houden* ‘to take into account’. This is a common feature of CS in general and it supports the idea of collocations being units in the mental lexicon. This is not restricted to idioms (cf. Levelt, 1989: 187); rather, any frequent co-occurrence of lexical items creates and strengthens collocational ties between them (Backus, 1996b: 225 ff.). Idiomatic collocations like the above examples will be accessed as one lexical unit; in other cases, ‘one element of a collocation may be selected and subsequently trigger the other elements, because of their collocational entrenchment’ (Backus, 1996b: 126). The latter procedure may account for the embedding of non-idiomatic adjective-noun combinations like *andere interesses* ‘other interests’, and complement-verb combinations like *wedstrijd voetballen* ‘to play a soccer game’. Collocational ties also account for the observation that embedded Dutch verbs select the same Dutch preposition they subcategorise for in monolingual utterances, and for the fact that embedded Dutch adverbs and predicative adjectives are (almost) exclusively modified by a Dutch adverb. An adverb-adjective combination like *heel aardig* ‘very kind’ may be viewed as a collocation, though not an idiomatic one.

If the existence of collocational ties between lexical units in the mental lexicon accounts for the co-occurrence of EL words, the total absence of such ties may

perhaps explain the observed constraints on the co-occurrence of ML and EL lexical items.

11.3.2 The Modification Restrictions

The modification restrictions concern the formal relationship between attributive adjectives and nouns, and between modifying adverbs and the adverbs and adjectives they modify. In the case of attributive adjectives, various studies on CS mention that they rarely occur as embedded forms (cf. Chapter 10, p. 362). Recall that in this respect, attributive adjectives contrast with predicative adjectives as the latter belong to the most common insertion types. Concerning adverbs, Lehtinen (1966: 175) notices that limiting (i.e. degree) adverbs are not inserted in her Finnish/English data. The MA/Dutch data confirm this: the embedded Dutch adverbs are either manner adverbs modifying verbs, or they have modal or discourse sequencing functions; they do not modify ML adjectives or adverbs. In addition to these restrictions that have been observed for MA/Dutch as well as for other CS varieties, the study of MA/Dutch revealed that embedded Dutch nouns, adjectives and adverbs tend not to be modified by MA (ML) attributive adjectives and adverbs, respectively. Whether this latter observation can be generalised to CS with other language pairs still needs to be established, but I assume the phenomenon will turn out not to be restricted to MA/Dutch.

The following examines the role played by collocational ties in an account of the regularities pertaining to the CS behaviour of adjectives and modifying adverbs. I will be talking mostly about adjectives, but much of the following reasoning also applies to the modification relationship between adverbs and adjectives or other adverbs.

Adverbs

Adverbs are less commented on in the CS literature than adjectives. Lehtinen (1966) notes that limiting (i.e. degree) adverbs are not inserted. She attributes this to their function morpheme-like status, since after all degree adverbs constitute a small and rather closed class (1966: 175). This is a possible explanation. EL degree adverbs are quite common, though, as modifiers of EL predicative adjectives or of other EL adverbs, and in this respect they differ from (other) free form function morphemes. (EL function morphemes that accompany EL content morpheme heads are typically affixes, cf. section 1.2.1 in Chapter 2.)

Adjectives

The dearth of EL attributive adjectives can be understood if we postulate that collocational ties are essential for the selection of attributive adjectives. A substantial amount of the attributive adjectives that occur in conversations are accessed as part

of a collocational unit.¹¹ We may further conjecture that in the process of speech production, the head noun is selected first, because it is more autonomous and because it more directly reflects the concept the speaker wants to convey (see Backus' (1996b) Specificity Continuum discussed above). If collocational ties between the attributive adjective and the head noun are a prerequisite for the selection of adjectival modifiers, EL adjectives have little chance of modifying ML nouns as they lack the collocational ties through which they can be activated. This 'collocation hypothesis' explains the divergent CS patterns for attributive and predicative adjectives. The latter are frequently inserted since they are not called upon as part of a collocation; to the contrary, the predicate contains new information on the (Subject) NP.

Now the analysis of the MA/Dutch data within the framework of the MSA yielded a new observation, namely that the modification of embedded nouns by ML adjectives is equally rare. The postulate that collocational ties between the head noun and the attributive adjective are essential for the selection of the latter could account for this part of the modification restrictions as well. Here also, the adjective would simply not be selected because nothing associates it with the head noun. At this point I do not know for certain whether the observation that EL nouns are rarely modified by ML adjectives applies to CS with other language pairs, but I would predict that it does. If so, this would constitute further support for the idea that the lack of collocational ties between lexical units causes the rarity of singly inserted attributive adjectives. In order for this type of explanation to acquire a solid basis, we would need independent evidence for the postulate that collocational ties are necessary, or, in a weaker form, play an important facilitating role in the selection of attributive adjectives.

Adjectives in Repetition Contexts

There are discourse contexts where the existence of collocational ties is less important. The clearest example of this occurs when part of the collocation is repeated for rhetoric purposes. Repetition may cause the repeated element to appear in a 'new' environment, far from units with which it shares collocational ties. This is illustrated in (26).

- (26) ma ta-ye-bği-w-š đyuṛ mwessx-in langs een straat mwessx-a
 NEG ASP-3-want-PL-NEG house·PL dirty-PL along a street dirty-F
 "They don't want dirty houses along a dirty street."
 (Hayat, ex. (9) in Ch. 9).

¹¹ An indication of this is the often idiomatic use of seemingly common adjectives. So in English, for example, *heavy*, besides referring to weight, denotes intensity in numerous expressions (e.g. with drinker, storm, odour, petting, gunfire). If we were to translate these expressions into, say, French, we would need many different words to express the intensity, and the idiomaticity of these uses of *heavy* then becomes apparent.

In (26) we do find a MA attributive adjective modifying a Dutch noun in a Dutch matrix constituent. The repetition of the MA adjective *mwessex* “dirty” emphasises the idea of dirtiness. Since MA insertions are so rare, we may safely assume that there are no collocational ties between MA *mwessex* and Dutch *straat* “street”. The former is re-selected for rhetorical purposes, and the latter for its referential meaning.¹² Being a repetition, the adjective *mwesxa* was probably selected prior to the head noun in this case.

Thus, repetition is a strategy in its own right which entails the possible juxtaposition of elements that are not normally found juxtaposed. The latter effect, however, is apparent only in bilingual texts that show clear co-occurrence restrictions on certain juxtapositions. For adjectives in CS this means that the constraint on modification can be overcome, either if the adjective itself is repeated as in (26), or when the modified noun is repeated as in (29), (30) and (33) below. It so happens that all counter-examples to the modification restrictions involve repetition. I will return to the use of repetition in section 4. In that section, I will also explain why it appears that the co-occurrence restriction for content words and modifiers from different languages holds only for CS with Moroccan Arabic as the ML.

Avoidance Strategies

There are some instances of CS that give the impression that speakers use strategies to avoid the modification of a Dutch EL noun by means of a MA adjective. These instances, discussed in Chapter 5, are repeated here as (27) and (28). In the first example Warda uses a MA relative clause consisting of no more than a resumptive pronoun and the MA adjective *ṣġira* “small” in a predicative position; in the second example Samir makes use of a roundabout genitive construction to express “these Dutch workers”.

- (27) ḥna sakn-in f dorp, hiya ṣġir-a
 1PL live-PART-PL in village 3F small-F
 “we live in a village, which is small” (Warda, ex. (84) in Ch. 5)

- (28) had l-arbeider-s dya l-huḷandi-yin
 DEM DEF-worker-PL of DEF-Dutch-PL
 “these Dutch workers” (Samir, ex. (85) in Ch. 5)

If it can be established that bilingual speakers use avoidance strategies to prevent EL nouns from being modified by an ML adjective, this means that there is a genuine modification constraint. After all, MA lemmas have been accessed for the expression

¹² Since the talk is about housing in the Netherlands, it may be that *straat* is selected rather than MA *zenqa* because streets in Morocco and in the Netherlands are different; it is also possible that the selection of the relatively specific locative preposition *langs* triggered a Dutch NP.

of the concepts SMALL in (27) and DUTCH in (28), despite the head nouns being in the EL, but the respondents ‘preferred’ not to use an ML attributive adjective. In other words, examples (27) and (28) suggest that the absence of collocational ties between the Dutch head noun and items in the MA lexicon does not prevent the activation of modifying concepts in MA, but pertains to the adjectival modification relationship per se.

Conclusion

The idea that collocational ties are a crucial facilitating factor for the activation of attributive adjectives is a possible explanation for several CS patterns involving adjectives. It could explain why attributive adjectives are not inserted, and why EL nouns are not modified by ML adjectives either. It would very adequately explain the differential behaviour of attributive and predicative adjectives. The ‘collocation hypothesis’ is further supported by the effect of repetition. Repetition is demonstrably the cause of the counter-examples to the co-occurrence restrictions on EL content word heads and ML modifiers and vice versa. The repetition effect, in its turn, can be explained by the idea that repetition overrules the impact of collocational ties. There are some examples showing that bilingual speakers do access ML lemmas to modify the embedded noun, while avoiding the modification relationship of the attributive adjective. This suggests that it is really the syntactic relationship between attributive adjective and head noun that is affected by collocational ties.

All things considered, the collocation hypothesis performs better than word order differences in explaining the CS behaviour of adjectives (Ch. 1, p. 17). The collocation hypothesis can be analogously applied to account for very similar facts pertaining to the CS behaviour of certain modifying adverbs, especially degree adverbs.

11.3.3 The Complementation Restriction

The description of MA/Dutch CS revealed that embedded Dutch verbs take only Dutch nouns as their lexical complements. This holds for both Direct Object complements and prepositional complements of embedded Dutch verbs. This is also probably a recurrent feature of CS in general. The Turkish/Dutch data studied by Backus (1992, 1996b) appear to be very similar to MA/Dutch: Dutch verbs are inserted in a Turkish periphrastic *do*-construction with the verb *yap-* and here too, the lexical complements of the embedded verbs are in Dutch whereas pronominal complements are exclusively in Turkish (Backus, personal communication; cf. Chapter 6, p. 264). Just as the modification restrictions are associated with the insertion of adjective-noun collocations and collocations of modifying adverbs and other adverbs and adjectives, the explanation for the complementation restriction may be related to the occurrence of embedded Verb-Object collocations. In any case, embedded Verb-Object collocations are found to be common in various language pairs; Backus (1996b: 278-80) lists a number of cases, and see also Boumans (1996).

True, this does not imply that EL verbs in these language pairs tend not to take ML lexical complements, as is the case for MA/Dutch. However, since the modification restrictions and the connection with EL adjective-noun combinations recur in several language pairs, I would expect the same in the case of the complementation restriction in connection with EL object-verb combinations. This remains to be investigated before a firm conclusion can be drawn.

So far the complementation restriction has been observed for embedded verbs in periphrastic *do*-constructions in MA/Dutch and Turkish/Dutch, and the phenomenon may be germane to this type of verb embedding only. In MA/French CS, for instance, embedded French verbs are inflected by means of attaching Arabic verbal affixes to the French verb stem, and there is no trace of the complementation restriction in this CS variety. In any case, the fact that the same observation is made for Turkish/Dutch makes it unlikely that the explanation lies in language-specific features of Moroccan Arabic and Dutch, such as the different ‘basic’ word orders in these languages (Verb-Object in MA; Object-Verb in Dutch), or the fact that the periphrastic “do”-construction is uncommon in monolingual MA.

Like the modification restrictions, the complementation restriction can be understood if we assume that the complementation relationship between verb and lexical complement typically involves collocational ties. Then, in a Verb-Object collocation like Dutch *wedstrijd voetballen* “play a soccer game” or *technieken toepassen* “apply techniques”, it can be argued that the lexical object is the more specific element, and the verb the less specific element. The verb and the object noun are also distinguished in terms of autonomy, which Backus subsumes under his definition of specificity: nouns can stand on their own, while verbs require one or more arguments to be specified (Backus, 1996b: 122). According to Backus’ Specificity Continuum, the most specific lexical unit in a collocation is likely to be accessed first, and subsequently triggers the less specific part(s) (Backus, 1996b: 126). The complementation restriction then results from an implicational hierarchy: insertion of the less specific element implies that the more specific element is inserted too. So we either find the insertion to be restricted to the most specific part of the collocation (e.g. *technieken* “techniques”), or the insertion is expanded to include the less specific verb (*technieken toepassen* “apply techniques”). However, it is unlikely that the more specific part would be realised in the ML, while the less specific part is embedded from the EL. In this way we arrive at the insertion patterns that were found in the MA/Dutch corpus: embedded Dutch verbs either have a highly specific embedded Dutch complement, or they have a MA complement situated at the low end of the specificity continuum, such as a pronominal complement. Or, of course, no complement is specified for the verb at all, as in the case of intransitive verbs. In each case, the embedded material comprises the more specific, new, and focused information in the clause, whether this is the verb itself, or its complement.

This explanation is analogous to the explanation of the modification restrictions: embedding of the more specific element (the head noun or the complement noun) is a precondition for the insertion of the less specific part of the collocation (the

attributive adjective or the verb). The analogy is only partial, though: ML verbs do occur with lexical EL Direct Object and other complements, but ML attributive adjectives do not modify EL nouns. So the selection of the attributive adjective depends more on the head noun than the activation of the verb depends on the lexical complement. This is because the verb may also occur without a lexical complement, whereas the modifier does not occur without the head. Furthermore, the verb has collocational ties with the inflectional categories (Tense, Aspect, et cetera) which, by definition, derive from the ML.

That collocational ties play such an important role in the selection of lexical complements is somewhat unexpected since the embedded verb is on average quite free in the selection of its complement, setting aside the real idioms. But the data bear evidence that something in the relationship between verbs and lexical complements, and not the complementation relationship itself, blocks MA nouns from occurring as the complement of an embedded Dutch verb: where applicable, MA clauses and pronouns do occur freely as complements, and there are even some instances of MA indefinite and demonstrative pronouns (*ši haža* “something”, *hadak* “this”).

11.3.4 Disappearance of co-occurrence restrictions

The explanation I propose for the attested co-occurrence restrictions is that they result from the lack of collocational ties which act as a facilitating factor in the selection of adjectival and adverbial modifiers and of lexical complements of embedded verbs. If this proves to be the case, and if collocational ties between content words result from their frequent co-occurrence, we may expect the co-occurrence restrictions in CS to disappear over time when counter-examples become increasingly frequent. Counter-examples do occur in MA/Dutch, notably in connection with repetition.

The ‘collocation hypothesis’ therefore engenders the prediction that co-occurrence restrictions are typical of the relatively new and unstable CS varieties; in more conventionalised CS varieties the co-occurrence restrictions will have disappeared or be less pronounced. A first comparison of the MA/Dutch data and Algerian Arabic/French data from Algiers (Boumans & Caubet, *forthc.*) seems to confirm such a relationship. In Algeria, where Arabic/French CS has a long history and where it is a feature of nearly all casual conversation, there is little evidence for the existence of co-occurrence restrictions on the combination of Arabic and French content words. Embedded French verbs occur freely with Arabic lexical complements. The insertion of French attributive adjectives and the adjectival modification of embedded French nouns is infrequent in Algerian Arabic/French CS as well, but not as rare as in MA/Dutch.

11.4 Repetition

Repetition as an independent mechanism is relevant to CS patterns in three different ways. Firstly, repetition as a means of creating textual cohesion can enhance the insertion of content words. Secondly, there is evidence that the same repetition overrules the working of the modification restrictions and perhaps the complementation restriction discussed in the previous section. Thirdly, the repetition of constituents may cause layered insertion.

11.4.1 Textual cohesion

In the discussion of a topic speakers stick to the original terminology, whether this was first introduced by themselves or by one of their interlocutors. This repetition contributes to the cohesion of a text, together with other grammatical and lexical mechanisms (Haliday & Hasan, 1976: 274-92). In addition to enhancing textual cohesion in a general, unmarked sense, repetition is sometimes used for special rhetoric and discourse organising effects. In bilingual conversations such repetition causes codeswitching when interlocutors keep to the terminology set at an earlier point in the discussion while shifting to another (matrix) language. In addition, embedded content words (from the Superimposed Language) tend to be repeated as embedded forms as the conversation continues in the same matrix language (De Rooij, 1996: 174-96). It has not yet been established whether embedded words have a higher propensity to be repeated than ML words, but if so, this might be related to the fact that EL words are typically highly specific. The MA/Dutch data further show that repetition is often involved in atypical CS patterns, viz. the insertion of MA content words in Dutch matrices, counter-examples to the modification restrictions, and layered insertion. This makes repetition a key factor accounting for many counter-examples to observed regularities.

Dutch content words in Moroccan Arabic

In (124) of Chapter 6 we saw an example of how repetition triggers the insertion of a Dutch content word in a MA matrix clause. In that example, Samir introduces the Dutch term *opvallen* ‘to be noticed’ into a conversation with Layla and Fatima on wearing the Islamic headscarf in the Dutch social context. Fatima asks Samir to explain this word to her and subsequently she uses this word herself as an embedded form in a MA clause, even though she would have been perfectly able to express her point in monolingual MA. Moreover, as it turns out, she has not entirely captured the meaning and the use of the Dutch term. In this example the repetition of *opvallen* only leads to some additional instances of CS, rather than to a new CS pattern. We may speculate, however, that repetition in the form of imitating another interlocutor (as in this example) plays a role in spreading particular CS patterns like the *dar* plus infinitive construction in the bilingual speech community.

Moroccan Arabic content words in Dutch

In Chapter 9, I demonstrated that repetition plays a central role for MA insertions in Dutch matrices. About half of all attested MA nouns and adjectives embedded in Dutch matrix constituents turned out to be repeated after having occurred in a monolingual MA clause in the immediately preceding part of the conversation. Furthermore, these insertions are clearly distinguished from the other half of the embedded MA content words: the latter, all of them nouns, refer to culturally specific Moroccan or Islamic concepts while repeated items are very common lexical items. So here we see repetition as a mechanism that generates an unusual insertion pattern. An example of a MA insertion caused by repetition is reproduced below.

(29) S waš ʕend-ek bit waḥd-a u ʕafi hna-ya?
 Q at-2SG room one-F and that's all here-EMPH
 “Do you have just one room here?”

J iyeh, hiya hadi
 yes 3F DEM
 “Yeah, that's it.”

S (..) u huwa ʕend-u *slaapkamer*, *en nog een ander-e* bit, *douche*
 and 3M at-3M bedroom and yet a other-AGR room shower
 “(..) So he [i.e. your housemate] has a bedroom, also another room, and
 a shower..” (Samir and Jamal, ex. (7) in Ch. 9)

Note in passing that in order for repetition to enhance textual cohesion, the repeated item need not necessarily be co-referential with its earlier occurrence; in fact in the above example it is expressly indicated that the two instances of *bit* refer to different rooms (cf. Halliday & Hasan, 1976: 282-4). Possibly repetition causes as many Dutch insertions in MA as vice versa. But it is a conspicuous and relatively important factor in CS with Dutch as the matrix language, as evidenced by the large amount of Dutch insertions in MA clauses caused by repetitions, while other factors related to the Superimposed/Community Language dichotomy generate only few MA insertions.

11.4.2 Counter-examples to the modification restrictions

In section 3.2 above I discussed repetition in connection with the modification restrictions. Recall that the modification restrictions refer to the observed tendency to avoid certain modification relationships involving an ML head (noun, adjective or adverb) and an EL content word modifier (attributive adjective, or adverb), and vice versa, an EL head with an ML modifier. Now it turns out that in nearly all counter-examples to this constraint in MA/Dutch CS, either the modifier or the head noun or adjective is a repeated item, irrespective of whether MA or Dutch is the ML. Concerning CS with MA as the ML, reconsider some examples from Chapter 5. In

that chapter (section 1.10) it was noted that embedded Dutch nouns are not modified by a MA adjective; in all three counter-examples to this rule either the Dutch noun or the MA adjective was a repeated item. Of these three, one is reproduced below. Also, I counted only one exception to the rule that embedded Dutch (predicative) adjectives are never modified by a MA degree adverb. (31) is a more complex example, because when the Dutch adjective *oud* occurs with the MA adverbs *šwiya* and *bezzaf*, it is not repeated literally, but as the translation of the MA word *kbir*.

- (30) b *š-šehh* *huwa racist, racist kbir hadak*
 with DEF-reality 3M racist racist big DEM
 “But he is a racist, a big racist he is.” (Younes, ex. (83) in Ch. 5)

- (31) *n-kun kbir, n-kun gaʕ kbir, šwiya oud, oud bezzaf*
 1-be old 1-be completely old a-little old old very
 “I’ll be old, very old, a little old, very old.”
 (Abdelkrim, ex. (106) in Ch. 5)

Apparently the repetition effect overrules the modification restrictions in CS with MA as the ML. When Dutch is the ML, at first sight the modification restrictions do not seem to apply at all. Of the total of 12 embedded MA nouns in the corpus (excluding the place-name *hulanda* “Holland”), three are modified by a Dutch ML adjective; *andere bit* “other room” in (29) exemplifies this. Compare this to the Dutch nouns in MA matrices: there are hundreds of these in the text corpus and yet no more than three of them are modified by a MA adjective. Also the scarcity of single attributive adjectives among the Dutch insertions in MA matrices is unparalleled when Dutch is the ML: true, (11) is the only example of a single MA attributive adjective in a Dutch nominal constituent, but this is not surprising, considering the overall paucity of MA insertions. To the examples with mixed noun-adjective combinations, we can add the only embedded MA predicative adjective, which happens to be modified by a Dutch adverb; consider *d šifa* “weak”, the lexical head of the mixed adjectival constituent *zó d šifa* “that weak”, cited here.

- (32) *ma dšif-a-š. zó dšif-a is ze particle*
 NEG weak-F-NEG that·much weak-F is she not
 “She’s not weak. She’s not all that weak.” (Fatima, ex (10) in Ch. 9)

We might conclude that the modification restrictions hold only when the ML is Moroccan Arabic, but not when the ML is Dutch. However, in each and every instance the mixed modifier-head combination results from repetition. Repetition, I argue, is a mechanism which overrules the modification restrictions. What seems at first sight to be a difference between the two matrix languages, is perhaps really the side-effect of repetition on CS patterns. The impression of asymmetric co-occurrence restrictions arises because repetition generates many of the MA embedded

content words per se, irrespective of whether they engage in a modifier-head relationship or not. Therefore, it is difficult to assess whether the modification restrictions apply when Dutch is the ML. Evidence should come from embedded MA nouns which are selected for their referential meaning, that is, the highly specific Moroccan and Islamic concepts discussed in section 1.1 of Chapter 9. There is only one counter-example to the modification restrictions and here the inserted noun happens to be repeated too: the MA noun *žellaba* ‘jellaba’ first occurs without the ML adjective in the first line of (33), and is then modified by the Dutch adjective *lange* ‘long’ in the second line.

- (33) *er is één meisje of één man die met een baard en met een žellaba,*
 there is one girl or one man who with a beard and with a jellaba

lang-e žellaba bedoel ik, ik bedoel ja, marokkaanse eh kleren,
 long-AGR jellaba mean I I mean yes Moroccan er clothes
 “(Suppose) there’s just one girl (with a headscarf) or one man with a beard
 and a jellaba, I mean a long jellaba, I mean, well, Moroccan er dress, (he’ll
 attract the people’s attention).” (Fatima, ex (6) in Ch. 9)

Why does repetition override the modification restrictions? The answer is of course related to the explanation of this constraint itself. In section 3.2 above I proposed to account for these co-occurrence restrictions by postulating that collocational ties between the modifier and the head are a prerequisite for the selection of the modifier. The modification restrictions would then result from the absence of collocational ties between the modifier and the head if these are from different languages. (An analogous explanation for the complementation restriction was advanced in section 3.3.) These restrictions can be overcome in repetition contexts because one of the functions of repetition is exactly this: to present an ‘old’ element in a ‘new’ textual environment. Where the modifier is reiterated, a new object is introduced which shares the same property as the one first mentioned. This was demonstrated in (11) above: the repetition of the MA adjective *mwessex* ‘dirty’ renders a series of dirty objects (“dirty houses in a dirty street”), which creates the rhetoric effect of “everything is dirty”. Likewise, the head word can be repeated with the purpose of adding a new feature to it. In (33), for instance, *žellaba* is reiterated in order to add the feature “long”: “a long jellaba”.

Counter-examples to the complementation restriction

The question may arise as to whether repetition also generates counter-examples to the complementation restriction. No instance of this was found in the MA/Dutch data, but I would expect a Moroccan Arabic NP to occur as the complement of an embedded Dutch verb if either the NP or the embedded verb would be repeated. It should be kept in mind, however, that NPs, unlike nouns and adjectives, usually recur

as pronominal forms when reiterated. This is what happens in (34): *l-werqa* “the paper” recurs as the object suffix *-ha*.

- (34) $\text{\textcircled{S}emmer-t-ha, \text{\textcircled{S}emmer-t } l-werqa, der-t-l-ha } invull-en$
 fill·out-1SG-3F fill·out-1SG DEF-paper do-1SG-to-3F fill·out-INF
 “I filled it out, I filled out the form, I filled it out.”
 (Mustafa, ex (13) in Ch. 6)

11.4.3 Layered insertion

Interestingly, when the repetition of an element leads to CS, this often coincides with layered insertion. In the third line of (29), for instance, the MA noun *bit* “room” is part of the Dutch NP *nog een andere bit* “yet another room” which functions as the complement of MA *send-u* “he has” in a MA matrix clause. Relatively many of the embedded MA content words resurface upon repetition in a Dutch NP which is itself part of a MA matrix clause. As yet I have no explanation for this.

Another way in which repetition and layered insertion coincide occurs when a constituent rather than a content word is reiterated, becoming an EL constituent in a matrix clause of the other language. If the repeated constituent contains an EL content word, its repetition in the context of the other language involved generates an instance of layered insertion. This phenomenon has already been discussed in Chapter 2; Eliasson’s (1995) Maori/English example is repeated here for convenience. Notice the Maori NP *te hearse*, which consists of the Maori definite article *te* and the English noun *hearse*. The Maori NP first occurs as part of a Maori clause in line 2, and is repeated as part of the English clause in line 3.

- (35) *That’s right, i te haere mātau ki Mātauri, ā, ka pāhi mai te .. te .. te .. te ..*
te hearse, e tū ana mātau i tō mea raka, i tō Hōhepa rā, nē, te pick up i a
Hōhepa. Ana, ka karanga atu, “Gee, what’s te hearse over there?” Nā, ka
karanga mai ētahi, “Ana, ko Haki, kei te whakahoki mai i Whakatāne.”

“That’s right, we were journeying to Matauri and the .. the .. the .. the .. the hearse went by, while we were assembled at what’s his name’s place, the home of Hohepa. Aye, we were picking Hohepa up. I called out “Gee, what’s the hearse over there?” Then some people replied, “It is Haki, being taken home from Whakatane.” Maori/English (Eliasson, 1995: 51; ex. (42) in Ch. 2)

Due to the repetition in this example, it becomes plausible that the Maori NP *te hearse* is inserted in the English finite clause, rather than the Maori function morpheme *te*. If the Maori article were a singly embedded morpheme, this would constitute a highly unusual insertion type. The repetition of the mixed NP for which Maori is the ML constitutes an extra argument in favour of the layered insertion analysis for the

English finite clause in line 3. I take this as supportive evidence for the possibility of layered insertion in general, even if no repetition is involved.

11.5 Summary

In this chapter I proposed a sociolinguistic dichotomy based on the relative status of the languages involved in CS, i.e. the Community Language and the Superimposed Language. Because this dichotomy is logically independent from the Matrix Language/Embedded Language dichotomy, we can investigate how patterns of insertional CS correlate with the social status of the ML. The Community Language is the prototypical matrix language in which content words from the Superimposed Language are embedded. Community Language insertions in Superimposed Language do occur but tend to be different in nature: they are often nominal and prepositional constituents, they tend to occur in clause-initial position and are mostly not motivated by the selection of semantically specific vocabulary items from the Community Language. However, in some bilingual communities the languages have a more equal status, and this probably results in more symmetric codeswitching patterns. In addition, language contact situations in which the Community/Superimposed Language dichotomy is obvious may be subdivided into various types, each correlating with typical insertion patterns.

Section 2 offered a closer look at Community Language insertions. It was shown that the MA insertions in Dutch matrices are typical of Community Language insertions generally. This kind of insertion is often represented by nominal and prepositional constituents and, unlike embedded constituents from the Superimposed Language, embedded constituents from the Community Language can freely include language-specific function morphemes like pronouns, determiners and case markers. This discrepancy can be understood if we consider embedded Community Language constituents as a continuation of the former 'default' matrix language role of the Community Language, albeit at a very local level. Subsequently, I elaborated on the proclivity of Community Language elements to take the clause-initial position either within or external to Superimposed Language finite clauses. I suggested three alternative explanations for this observation: firstly, the tendency for Superimposed language elements to be highly specific, and hence to occur in focused position; secondly, a tendency towards the encoding of topical information in the Community Language as a kind of discourse marking device; and thirdly, a simple tendency for speakers to switch to the Superimposed Language after the first word or constituents, possibly using the Community Language opening entry as a marker of ethnic identity. Neither of these explanations necessarily excludes the others, but more research is required in order to determine whether there is actually a convergence of several factors or one fundamental explanation.

The co-occurrence restrictions constitute a separate, intriguing topic. They refer to the observation that some logically probable insertion patterns are conspicuously

absent. More specifically, in MA/Dutch CS, embedded Dutch verbs do not occur with MA lexical complements. The same observation was made for Turkish/Dutch. In addition, in MA/Dutch, mixed combinations of nouns and modifying adjectives, and of adjectives and adverbs modified by adverbs, are evidently avoided. This is partially confirmed by studies of CS with other language pairs, where the insertion of attributive adjectives is often reported to be uncommon. I refer to these co-occurrence restrictions as the complementation and modification restrictions respectively. The nature of these restrictions and the extent to which the observations on MA/Dutch can be generalised to other CS data need to be investigated. As a first step towards the explanation of the restrictions, I pointed out that they involve the same formal relationships as the collocations of content words which are often found as embedded forms: noun-adjective and verb-object collocations. The co-occurrence restrictions could be explained if we postulate that collocational ties are a requirement for the co-occurrence of lexical items in these relationships, or indeed, if the existence of collocational ties can be shown to be an important facilitating factor. Repetition of content words is a mechanism that overrules the co-occurrence restrictions. Repeated content words may enter into a modification or complementation relationship, despite the existing restrictions. If we assume that collocational ties emerge from the frequent co-occurrence of lexical items, we may assume that during the course of time, co-occurrence restrictions will attenuate or disappear between words that are being used in the same speech variety.

Repetition is a discourse mechanism in its own right which turns out to be responsible for many of the less common insertion patterns. It was shown that nearly all counter-examples to the modification restriction in the MA/Dutch corpus involve the repetition of either the modifier or the modified head word. Repetition of content words as a means of enhancing cohesion in the discourse generates intrasentential CS when more than one language is used during the discourse. Repetition of entire complex constituents may lead to constituent insertion, if the constituent is repeated as part of a matrix clause (or higher order constituent) in the other language. If, in addition, the repeated constituent already contains an embedded element, its repetition may even cause layered insertion.

Outlooks

Much of the present day research is concerned with psycholinguistic and sociolinguistic explanations for attested regularities in CS behaviour. The choice of the descriptive paradigm is crucial as it inevitably influences which patterns are observed. No data description is theoretically neutral. With respect to matrix language approaches to CS, the definition of the matrix language determines what the categories of embedded material will look like. Various topics could be further investigated from a crosslinguistic perspective, using the Monolingual Structure Approach for the identification and classification of insertion types in a uniform manner, and using the Community Language/Superimposed Language dichotomy as a sociolinguistic parameter.

Many aspects of codeswitching still call for an explanation; I will mention only a few. With respect to discourse marking devices, for instance, it has often been noted that they occur in the context of a language other than their source language, however this matter needs more clarification. The Community versus Superimposed Language dichotomy has proven to be more useful here than the distinction between matrix and embedded languages. So we may investigate what kinds of discourse markers from the Community Language tend to occur in Superimposed Language contexts, and vice versa. Another unsolved question concerns the word order of embedded attributive adjectives. This is not consistently predicted by the matrix language in itself, but perhaps a combination of variables including the Community Language/Superimposed Language dichotomy will perform with more success. The insertion of constituents is generally restricted. In this study it has been shown that constituent insertion is partly dependent on whether the Community Language or the Superimposed Language functions as the matrix language on the finite clause level. While embedded NPs and PPs from the Community Language may freely contain function morphemes which have no counterpart in the Superimposed Language system, only certain types of EL constituents from the Superimposed Language regularly occur. Setting aside this difference, the possibilities for constituent insertion will differ from one language pair to another. It seems likely that more types of EL constituents (from the Superimposed Language) are possible when the ML and EL are closely related, genetically or areally, as a result of previous language contact. This is another topic for the research agenda.

I would like to draw special attention to the questions raised in Chapter 11. Firstly, how do various patterns of insertion correlate with types of bilingual communities, and, if such correlations can be established, how do we account for them? Secondly,

the characteristics of insertions from the Community Language should be further investigated and accounted for. In particular, one of the questions which still awaits a satisfactory answer is why such insertions tend to occur in clause-initial position. Thirdly, various matters concerning the co-occurrence restrictions observed for Moroccan Arabic/Dutch need to be clarified. It is not clear to what extent these restrictions are features of codeswitching generally. This aside, further explanations are called for. Are the co-occurrence restrictions indeed related to collocations, as I have suggested, or is there a better explanation? If co-occurrence restrictions and the occurrence of (embedded) collocations derive from a common cause, we need to investigate the role of collocational ties in accessing lemmas from the mental lexicon. Finally, the role of repetition in discourse deserves more attention, since it is involved in many instances of less common insertion patterns.

Notational Conventions

In the main text

When referring to language pairs in a CS context, the two languages are ordered according to the principle Community Language/Superimposed Language (cf. Chapter 11, section 1.1 for these terms), for instance Moroccan Arabic/Dutch, Spanish/English. The sociolinguistic dichotomy is used to contrast the two languages as this makes it possible to refer to the same corpus and the same bilingual context in a consistent way, even though the matrix language as defined in Chapter 2 may vary within the same corpus, and a different matrix language may exist at the constituent and at the sentence level.

Italics are used for stretches cited from the numbered examples, as when referring to *m ša-h* in (2) below. SMALL CAPITALS indicate core terminology and emphasis. Initial capitals are used for the words Subject, Object and Topic when these refer to syntactic or pragmatic functions.

The following abbreviations are used in the main text:

B	borrowing
CL	Community Language
CS	codeswitching
DO	Direct Object
EL	embedded language
IO	Indirect Object
MA	Moroccan Arabic
ML	matrix language
MLF	Matrix Language Frame
MSA	Monolingual Structure Approach
NP	nominal constituent
PP	prepositional constituent
SL	Superimposed Language
VP	verbal constituent

In numbered examples

In cited examples of mixed sentences, the words and morphemes from the economically and culturally Superimposed Language are italicised, irrespective of which language functions as the Matrix Language.

- (1) *puis après ta-y-mesh-u la cicatrice*
 then after ASP-3-wipe-pl DEF·F scar
 “Then later they wipe off the scar.”
 Moroccan Arabic/French (Slaoui, 1986, Annexe III: 14)
- (2) f t-tali il m' a emmené mʃa-h
 in DEF-end he me has taken with-3M
 “In the end he took me with him.”
 Moroccan Arabic/French (Bentahila & Davies, 1991: 384)

Round brackets () in the cited example mark text fragments that facilitate the interpretation of the example but are not considered in the discussion. Underscored spaces ___ indicate the position from which morphemes are missing according to matrix language grammar. Square brackets [] indicate missing morphemes in the glosses. In the English translations square brackets mark words that are added to make the translation interpretable or well-formed, but do not necessarily indicate missing morphemes in the example text.

In the glosses grammatical categories are indicated in small capitals, using the following abbreviations:

1,2,3	first, second, third person
AFFIRM	affirmative particle
AGR	agreement
ASP	aspect
C	common gender
COMP	complementizer
COMPAR	comparative
DEF	definite article
DEM	demonstrative
DIM	diminutive
EMPH	emphatic
EPIST	epistemic
EXIST	existential
F	feminine gender
GEN	genitive
IMP	imperative
INDEF	indefinite article
INF	infinitive

M	masculine gender
MP	medio-passive or reciprocal
NEG	negation
PL	plural
POSS	possessive
PRES	presentative
REFL	reflexive
REL	relative clause marker
SG	singular
TOP	topic

Dots are used in the glosses to indicate that there is no one-to-one correspondence of morphemes and glosses. For instance the Dutch definite article *de* is glossed as either DEF•C ‘definite common gender’ or DEF•PL ‘definite plural’.

Transcription of Moroccan Arabic

Unfortunately there is as yet no generally accepted orthography for Moroccan Arabic. The transcription system used in this study is a phonological one based on Otten’s (1983) Moroccan Arabic-Dutch dictionary except for some small changes. Different symbols are used for some phonemes. In particular, a dot underneath the symbol replaces the overstrike hyphens used in Otten (1983) for the pharyngealised consonants and the voiceless pharyngeal fricative (/ṭ ḍ ṣ ẓ ṛ ḷ ḃ/ and /ħ/ instead of /ṭ̣ ḍ̣ ṣ̣ ẓ̣ ṛ̣ ḷ̣ ḅ̇/ and /ħ̣/). The voiced uvular fricative is written /ġ/, the voiced pharyngeal fricative /ʕ/, glottal stop /ʔ/, instead of Otten’s /g/, /e/ and /ʔ/. Also, only five vocalic phonemes are distinguished here, leaving out the short vowels /ĩ/ and /ǣ/ used by Otten in the transcription of Standard Arabic words. Whenever examples in Moroccan or other varieties of Arabic are cited from published sources, the transcription is adapted to the system used for the examples from the Nijmegen corpus, with due regard for dialectal differences. The following table, based on Heath (1989: 17), presents the MA consonant inventory (*pharyng.* stands for ‘pharyngealised’).

	voiceless stop	voiced stop	nasal	voiceless fricative	voiced fricative	voiceless sibilant	voiced sibilant	lateral	rhotic	aspiration	semivowel
plain labial		b	m	f							w
pharyng. labial		ɸ	ɱ								
plain alveolar	t	d	n			s	z	l	r		
pharyng. alveolar	ṭ	ḍ				ṣ	ẓ	ḷ	ṛ		
palato-alveolar						š	ž				y
velar	k	g									
uvular		q		x	ġ						
pharyngeal				ħ	ʕ						
laryngeal	ʔ									h	

Moroccan Arabic consonants (after Heath, 1989: 17)

Five vocalic phonemes are distinguished: the full vowels /a u i/ and the unstable, often evanescent ‘short’ vowels /e/ and /ũ/. The phonetic realisation of the vowels is heavily dependent on the adjacent consonants.

Verb morphology

MA verbs have two conjugation paradigms, known as perfect and imperfect. The perfect paradigm is characterised by Subject agreement suffixes, except for the third person masculine, which can be said to have a zero suffix. The imperfect paradigm has agreement suffixes for all persons in addition to suffixes for all plural forms and feminine second person singular. For ease of reference the paradigms are given below with their corresponding glosses.

		imperfect		perfect	
SG	1	<i>ne-šreb</i>	1-drink	<i>šreb-t</i>	drink-1SG
	2M	<i>te-šreb</i>	2-drink	<i>šreb-ti</i>	drink-2SG
	2F	<i>t-šerb-i</i>	2-drink-F	<i>šreb-ti</i>	drink-2SG
	3M	<i>ye-šreb</i>	3-drink	<i>šreb</i>	drink
	3F	<i>te-šreb</i>	3F-drink	<i>šerb-et</i>	drink-3F
PL					
	1	<i>n-šerb-u</i>	1-drink-PL	<i>šreb-na</i>	drink-1PL
	2	<i>t-šerb-u</i>	2-drink-PL	<i>šreb-tu</i>	drink-2PL
	3	<i>y-šerb-u</i>	3-drink-PL	<i>šerb-u</i>	drink-3PL

The above paradigm represents the Atlantic Coast Koine form of MA as well as most other Moroccan dialects. The East Moroccan bedouin dialects spoken by some of the respondents (cf. Chapter 4) distinguish masculine and feminine second person singular perfect:

		perfect	
SG	1	<i>šreb-t</i>	drink-1SG
	2M	<i>šreb-t</i>	drink-2M
	2F	<i>šreb-ti</i>	drink-2F
	3M	<i>šreb</i>	drink
	3F	<i>šerb-at</i>	drink-3F

The Subject agreement in perfect verbal forms should not be confused with Direct and Indirect Object suffixes, which are listed below:

		Direct Object		Indirect Object		
		post-vocalic allomorph				
SG	1	<i>-ni</i>		<i>-l-i</i>	~	<i>li-ya</i>
	2	<i>-ek</i>	<i>-k</i>	<i>-l-ek</i>	~	<i>li-k</i>
	3M	<i>-u</i>	<i>-h</i>	<i>-l-u</i>	~	<i>li-h</i>
	3F	<i>-ha</i>		<i>-l-ha</i>	~	<i>li-ha</i>
PL						
	1	<i>-na</i>		<i>-l-na</i>	~	<i>li-na</i>
	2	<i>-kũm</i>		<i>-l-kũm</i>	~	<i>li-kũm</i>
	3	<i>-hũm</i>		<i>-l-hũm</i>	~	<i>li-hũm</i>

The East Moroccan dialect has the third person masculine singular DO suffix *-eh* instead of *-u*. Sometimes masculine and feminine gender are distinguished in the second person singular, in which case the forms are *-ek ~ -k* for masculine and *-ki* for feminine. The DO suffixes are the same as the pronominal complements of prepositions and the possessive suffixes, except for the first person singular, which is *-i* with the postvocalic allomorph *-ya* in these cases.

References

- ABBASSI, ABDELAZIZ. 1977. A sociolinguistic analysis of multilingualism in Morocco. PhD Thesis, University of Texas at Austin.
- ABOU, SELIM. 1962. *Le bilinguisme arabe-français au Liban*. Paris: Presses Universitaires de France.
- AITCHISON, JEAN. 1991. *Language chance: Progress or decay?* Cambridge: Cambridge University Press.
- ANDREWS, AVERY. 1985. The major functions of the noun phrase. In Shopen, 62-154.
- ANDROULAKIS, GEORGE. 1994. Towards a formal analysis of the Greek-French code-switching in Paris. In Philippaki-Warbuton et al., 357-64.
- ANNAMALAI, E. 1971. Lexical insertion in a mixed language. *Papers from the seventh regional meeting of the Chicago Linguistic Society*, 20-27. Chicago, IL: Chicago Linguistic Society.
- APPEL, RENÉ, and PIETER MUYSKEN. 1987. *Language contact and bilingualism*. London etc.: Edward Arnold.
- AQUILINA, JOSEPH. 1965. *Teach yourself Maltese*. Valletta: Hodder and Stoughton. [1987 reprint]
- ARTHUR, JO. 1995. Code-switching and collusion: classroom interaction in Botswana primary school. In *Summer School Code-Switching and Language Contact*, 1-15.
- AUER, PETER. 1984. On the meaning of conversational code-switching. *Interpretive sociolinguistics: migrants - children - migrant children*, ed. by Peter Auer & Aldo di Luzio, 87-108. (Ergebnisse und Methoden moderner Sprachwissenschaft 20.) Tübingen: Günter Narr.
- AZGHARI, YOUSSEF. 1994. Nederlandse ontleningen in het Marokkaans Arabisch. Master's Thesis, University of Nijmegen.
- BACKUS, AD. 1992. *Patterns of language mixing. A study in Turkish-Dutch bilingualism*. Wiesbaden: Harrassowitz.
- _____. 1996a. Het intergenerationele codewisseling-continuüm binnen de Turkse gemeenschap in Nederland. *Toegepaste Taalwetenschap in Artikelen*, 54(1).25-37.
- _____. 1996b. *Two in one. Bilingual speech of Turkish immigrants in the Netherlands*. (Studies in Multilingualism 1.) Tilburg: Tilburg University Press.
- _____. and LOUIS BOUMANS. 1996. Codewisseling en taalvariatie bij Turkse en Marokkaanse jongeren. *Etnische minderheden en wetenschappelijk onderzoek*,

- ed. by Henk Heeren, Patricia Vogel and Hans Werdmölder, 141-63. Meppel: Boom.
- BARHOUM, KHALIL. 1994. English-Arabic code-switching as a form of bilingual discourse. *Investigating Arabic: Linguistic, pedagogical and literary studies in honor of Ernest N. McCarus*, ed. by Raji M. Rammuny and Dilworth B. Parkinson, 95-109. Columbus, OH: Greyden Press.
- BASFAO, KACEM, and HINDE TAARJI (Eds.). 1994. *L'annuaire de l'émigration*. Maroc. Rabat: Annuaire de l'émigration.
- BAUTISTA, MARIA LOURDES S. 1975. A model of bilingual competence based on an analysis of Tagalog-English code-switching. *Philippine Journal of Linguistics* 6(1).51-89.
- . 1980. *The Filipino bilingual's linguistic competence: A model based on an analysis of Tagalog-English code-switching*. (Pacific Linguistics Series C-59.) Canberra: Australian National University.
- BEAUMONT, JEAN-CHARLES. 1987. L'emprunt en arabe marocain: facteurs socio-linguistiques d'intégration. *McGill Working Papers in Linguistics* 4(2). 53-88.
- BELAZI, HEDI M., EDWARD J. RUBIN and ALMEIDA J. TORIBIO. 1994. Code switching and X-bar theory: The functional head constraint. *Linguistic Inquiry* 25(2).221-37.
- BENTAHILA, ABDELALI and EIRLYS E. DAVIES. 1983. The syntax of Arabic-French code-switching. *Lingua* 59.301-30.
- . and EIRLYS E. DAVIES. 1991. Constraints on code-switching: a look beyond grammar. In *Papers for the Symposium on Code-Switching in Bilingual Studies*, vol. 2, 369-403.
- . and EIRLYS E. DAVIES. 1992. Code-switching and language dominance. *Cognitive processing in bilinguals*, ed. by Richard Jackson Harris, 443-58. (Advances in Psychology 83.) Amsterdam: North-Holland.
- . and EIRLYS E. DAVIES. 1994. Two languages, three varieties: codeswitching patterns of bilingual children. *The cross-linguistic study of bilingual development*, ed. by Guus Extra and Ludo Verhoeven, 113-28. Amsterdam etc.: North-Holland.
- BERK-SELIGSON, SUSAN. 1986. Linguistic constraints on intrasentential code-switching: a study of Spanish/Hebrew bilingualism. *Language in Society* 15.313-348.
- BLOM, JAN-PETTER and JOHN J. GUMPERZ. 1972. Social meaning in linguistic structures: code-switching in Norway. *Directions in Sociolinguistics. The ethnography of communication*, ed. by John J. Gumperz and Dell Hymes, 407-34. New York, NY: Holt, Rinehart & Winston.
- BOESCHOTEN, HENDRIK E. and LUDO TH. VERHOEVEN. 1987. Language-mixing in children's speech: Dutch language use in Turkish discourse. *Language Learning* 37(2).191-215.
- BOKAMBA, E.G. 1989. Are there syntactic constraints on code-mixing? *World Englishes* 8.277-93.

- BOLLE, JETTE. 1995. Mengelmoes: Sranan and Dutch language contact. In *Summer School Code-Switching and Language Contact*, 290-4.
- BORER, HAGIT. 1986. Introduction. *The Syntax of Pronominal Clitics* (Syntax and Semantics 19), ed. by Hagit Borer, 1-11. Orlando, FL (etc): Academic Press.
- BOUMANS, LOUIS. 1994. L'arabe marocain et le néerlandais en contact: quelques particularités syntaxiques. Paper read at Troisième Rencontre Internationale sur les Mouvements Migratoires Magrébins, Oujda, Morocco.
- _____. 1995a. The integration of verbs in codeswitching. In *Summer School Code-Switching and Language Contact*, 294-8.
- _____. 1995b. Le rôle des morphèmes de fonction dans le codeswitching. *Langues du Maroc. Aspects linguistiques dans une contexte minoritaire*, ed. by Petra Bos, 55-68. Tilburg: Tilburg University Press.
- _____. 1996. Embedding verbs and collocations in Moroccan Arabic/Dutch codeswitching. In Eid and Parkinson, 45-67.
- _____. 1997. Review of Heath 1989. *Bibliotheca Orientalis* 54(1/2).249-52.
- _____. forthcoming. Evidence from codeswitching for the organisation of the mental lexicon, with special reference to Moroccan Arabic. *Bilingualism and migration*, ed. by Guus Extra and Ludo Verhoeven, 00-00.
- _____ and DOMINIQUE CAUBET. forthcoming. A comparative study of codeswitching in Algeria and in the Netherlands. *Arabic as a minority language*, ed. by Jonathan Owens, 00-00.
- BROWN, BECKY. 1986. Cajun/English code-switching: a test of formal models. *Diversity and Diachrony* (Current Issues in Linguistic Theory 53), ed. by David Sankoff, 399-406. Amsterdam: John Benjamins.
- CAMILLERI, ANTOINETTE. 1996. Language values and identities: Codeswitching in secondary classrooms in Malta. *Linguistics and Education* 8(1).85-104.
- CAUBET, DOMINIQUE. 1993. *L'arabe marocain. I: Phonologie et morphosyntaxe, II: Syntaxe et catégories grammaticales, textes*. Paris/Louvain: Peeters.
- _____. n.d. [1995]. Enunciative particles in Moroccan Arabic: *ba'da* and *za'ma*. *Proceedings of the 2nd international conference of L'association internationale pour la dialectologie arabe*, 21-30. [Cambridge]: [University of Cambridge].
- _____. 1997. Transcripts of Md. Fellag's theatre shows; unpublished data.
- _____. 1998. Etude sociolinguistique des traits préhilaliens dans un dialecte en voie d'urbanisation à Fès. *Peuplement et arabisation au Maghreb occidental. Dialectologie et histoire*, ed. by Jordi Aguadé, Patrice Cressier and Angeles Vincente, 165-75. Madrid/Zaragoza: Casa de Velazquez.
- Centraal Bureau voor de Statistiek. 1986. *De leefsituatie van Turken en Marokkanen in Nederland. Deel 2: Kerncijfers* (CBS publicaties). 's Gravenhage: Staatsuitgeverij.
- CHOI, JAE OH. 1991. Korean-English code switching: switch-alpha and linguistic constraints. *Linguistics* 29.877-902.

- CLYNE, MICHAEL G. 1967. *Transference and triggering. Observations on the language assimilation of postwar German speaking migrants in Australia*. The Hague: Nijhoff.
- _____. 1971. German-English bilingualism and linguistic theory. *Dichtung, Sprache und Gesellschaft. Akten des 4. internationalen Germanistenkongresses*, ed. by V. Lange and H.G. Roloff, 503-11. Frankfurt: Athenäum.
- _____. 1972. *Perspectives on Language Contact. Based on a Study of German in Australia*. Melbourne: Hawthorn.
- COHEN, MARCEL. 1912. *Le parler arabe des juifs d'Alger*. Paris: H. Champion.
- COLIN, G.-S. 1945. L'arabe. *Initiation au Maroc*, 219-44. Paris: Vanoest.
- _____. 1993. *Le dictionnaire Colin d'Arabe Dialectal Marocain* (8 vols.). Rabat: Al Manahil- Ministère des Affaires Culturelles.
- COMRIE, BERNHARD. 1981. *Language Universals and Linguistic Typology*. Chicago, IL: Univ. of Chicago Press. [1989 edition]
- CORNIPS, LEONIE. 1994. De hardnekkige vooroordelen over de regionale doen+ infinitief-constructie. *Forum der Letteren* 35(4).282-94.
- DE ROOIJ, VINCENT A. .1996. *Cohesion through Contrast: Discourse structure in Shaba Swahili/French conversations*. (Studies in Language and Language Use 24) Amsterdam: IFOTT. [PhD Thesis, University of Amsterdam]
- DE RUITER, JAN JAAP. 1994. Hollandais et Marocains face à l'enseignement: données comparatives. In Basfao and Taarji, 519-21.
- DE VRIENDT, SERA, and PIET VAN DE CRAEN. 1993. Voegwoordelijke partikels en Frans/Nederlandse codewisseling in het Brussels. *De Brusselse rand. Colloquium 9-10 december 1993, Brussel*, ed. by Els Witte, 495-511. Brussel: VUB Press.
- DEETERS, GERHARD. 1926. Armenisch und Südkaukasisch; ein Beitrag zur Frage der Sprachmischung. *Caucasica* 3.57-82.
- DEUCHAR, MARGARET. 1995. Do two word utterances by bilingual children show evidence of code-switching? In *Summer School Code-Switching and Language Contact*, 39-44.
- DIEBOLD, A. RICHARD. 1963. Code-switching in Greek-English bilingual speech. *Report of the thirteenth annual round table meeting on linguistics and language studies*, ed. by Elizabeth D. Woodworth and Robert J. Di Pietro 53-9. (Georgetown Univ. Monograph Series on Languages and Linguistics 15.) Washington, DC: Georgetown University Press.
- DI SCIULLO, ANNE-MARIE, PIETER MUYSKEN and RAJENDRA SINGH. 1986. Government and codeswitching. *Linguistics* 220.1-24.
- DONALDSON, BRUCE. 1997. *Dutch: A comprehensive reference grammar*. London etc.: Routledge.
- DORON, EDIT. 1983. On a formal model of code-switching. *Texas linguistic forum* 22, ed. by Mary Dalrymple, Edit Doron, John Goggin, Beverley Goodman and John McCarthy, 35-59. Austin, TX: University of Austin, Department of Linguistics.

- DRAPEAU, LYNN. 1995. Code switching in caretaker speech and bilingual competence in a native village of northern Quebec. *International Journal of the Sociology of Language* 113:157-164.
- DURÁN, RICHARD P. (Ed.) *Latino language and communicative behaviour*. Norwood, NJ: Ablex.
- DURAND, OLIVIER. 1994. *Profilo di Arabo Marocchino. Varietà urbane centro-meridionali* (Studi Semitici, nuova serie 11.) Roma: Università La Sapienza.
- EID, MUSHIRA. 1985. On the function of pronouns in Egyptian Arabic. *Beyond the sentence: discourse and sentential form*, ed. by Jessica R. Wirth, 31-44. Ann Arbor: Karoma.
- _____. 1992a. Directionality in Arabic-English code-switching. In Rouchdy, 50-71.
- _____. 1992b. Pronouns, questions, and agreement. *Perspectives on Arabic linguistics IV*, ed. by Ellen Broselow, Mushira Eid and John McCarthy, 107-141. (Current Issues in Linguistic Theory 85.) Amsterdam/Philadelphia: Benjamins.
- _____. 1996. Things are not what they seem: pronoun doubling in bilingual and monolingual grammars. *Perspectives on Arabic linguistics VIII*, ed. by Mushira Eid, 7-29. (Current Issues in Linguistic Theory 134.) Amsterdam/Philadelphia: Benjamins.
- EID, MUSHIRA, and DILWORTH PARKINSON (Eds.) *Perspectives on Arabic Linguistics IX* (Current Issues in Linguistic Theory 141.) Amsterdam/Philadelphia: Benjamins.
- EL AÏSSATI, ABDERRAHMAN. 1996. Language loss among native speakers of Moroccan Arabic in the Netherlands. PhD Thesis, University of Nijmegen.
- ELIASSON, STIG. 1995. Grammatical and lexical switching in Maori-English 'grasshopper speech'. In *Summer School Code-Switching and Language Contact*, 45-57.
- EL-IDRISSI, ABDELJALIL. 1990. Verbes opérateurs et nominalisation en arabe marocain (parler de Fès). Thèse de Doctorat, Université de Jussieu (Paris VII).
- ENNIGER, WERNER, and JOACHIM RAITH. 1988. Varieties, variation, and convergence in the linguistic repertoire of the Old Order Amish in Kent County, Delaware. *Variation and Convergence. Studies in Social Dialectology*, ed. by Peter Auer and Aldo di Luzio, 260-93. (Soziolinguistik und Sprachkontakt 4.) Berlin/New York: Walter de Gruyter.
- ESPINOSA, AURELIO. 1917. Spanish mixture in New Mexico: the influence of the English language on New Mexican Spanish. *The Pacific Ocean in history*, ed. by H.M. Stephens and H.E. Bolton, 408-28. New York, NY: MacMillan. [Reprinted in Hernández-Chavez et al., 99-114.]
- EWING, ANNY. 1984. Polish-English code-switching: a clue to constituent structure and processing mechanisms. *Papers from the Twentieth Regional Meeting, Chicago 26-27 April 1984*, ed. by Joseph Drogo, Veena Mishra, and David Testen, 52-64. Chicago, IL: Chicago Linguistic Society.

- FAUSEL, ERICH. 1959. *Die deutschbrasilianische Sprachmischung*. Berlin: Erich Schmidt.
- . 1962. Hochdeutsch, Mundart und Mischsprache bei den Deutschen in Brasilien. *Wirkendes Wort* 12.210-7.
- FISHMAN, JOSHUA A. 1965. Who speaks what language to whom and when? *La linguistique* 2.67-88.
- FOLEY, WILLIAM A. 1986. *The Papuan Languages of New Guinea*. Cambridge: Cambridge University Press.
- and ROBERT D. VAN VALIN JR. 1985. Information packaging in the clause. In Shopen, 282-365.
- FORKEL, FRITZ. 1980. Die sprachliche Situation im heutigen Marokko. Eine soziolinguistische Untersuchung. Dissertation Universität Hamburg.
- GARDNER-CHLOROS, PENELOPE. 1991. *Language selection and switching in Strasbourg*. Oxford: Clarendon.
- GEERTS, G., J. HASEREYN, J. DE ROOIJ and M. VAN DEN TOORN (Eds.). 1984. *Algemene Nederlandse spraakkunst*. Groningen etc.: Wolters Noordhoff etc.
- GIBBONS, JOHN. 1987. *Code-mixing and code choice: A Hong Kong case study*. Clevedon: Multilingual Matters.
- GIESBERS, HERMAN. 1984. "Doe jij even lief spelen?" Notities over het perifrastisch doen. *Mededelingen NCDN* 19.57-76.
- . 1989. *Code-switching tussen dialect en standaardtaal*. Amsterdam: P.J. Meertens Instituut.
- GOULET, ROSALINA M. 1971. *English, Spanish and Tagalog: A study of grammatical, lexical and cultural interference*. (Philippine Journal of Linguistics Special Monograph.) Manila: Linguistic Society of the Philippines.
- GUMPERZ, JOHN J. 1964. Hindi-Punjabi code-switching in Delhi. *Proceedings of the Ninth International Congress of Linguists*, 1115-24.
- and EDUARDO HERNÁNDEZ-CHAVEZ. 1971. Cognitive aspects of bilingual communication. *Language use and social change*, ed. by W.H. Whiteley, 111-25. Oxford: Oxford University Press. [Reprinted in Hernández-Chavez et al., 154-164.]
- GÜNTHER, SUSANNE. 1993. "weil - man kann es ja wissenschaftlich untersuchen" - Diskurspragmatische Aspekte der Wortstellung in WEIL- Sätzen. *Linguistische Berichte* 143.37-59.
- HALLIDAY, MICHAEL A.K. and RUQAIYA HASAN. 1976. *Cohesion in English*. London: Longman.
- HALMARI, HELENA. 1993. Structural relations and Finnish-English code switching. *Linguistics* 31.1043-68.
- HARRELL, RICHARD S. 1962. *A short reference grammar of Moroccan Arabic*. Washington, DC: Georgetown University Press.
- HASSELMO, NILS. 1969. On diversity in American Swedish. *Svenska landsmål och svenskt folkliv* 92.53-72.

- _____. 1970. Code-switching and modes of speaking. *Texas studies in bilingualism*, ed. by Glenn G. Gilbert, 179-210. Berlin: de Gruyter.
- _____. 1972a. Språkväxling. *Språk och samhälle*, ed. by Bengt Loman, 152-82. Lund: Liber Läromedel.
- _____. 1972b. Code-switching as ordered selection. *Studies for Einar Haugen presented by friends and colleagues*, ed. by Evelyn Sherabon Firchow, Kaaren Grimstad, Nils Hasselmo and Wayne A. O'Neil, 261-80. The Hague/Paris: Mouton.
- _____. 1974. *Amerikasvenska. En bok om språkutvecklingen i svensk-Amerika*. (Skrifter utgivna av svenska språknämnden 51.) [Stockholm]: Esselte Studium.
- _____. 1975. Bilingualism as an emerging sociolinguistic system. *The Nordic languages and modern linguistics 2*, ed. by Karl-Hampus Dahlstedt, 245-66. Stockholm: Almqvist and Wiksell.
- _____. 1980. Spanish-English bilingualism in the US: comments on research during the 1970s. *The Nordic languages and modern linguistics*, ed. by Even Hovdhaugen, 157-69. Oslo: Universitetsförlaget.
- HAUGEN, EINAR. 1950. The analysis of linguistic borrowing. *Language* 26(2). 210-31.
- _____. 1953. *The Norwegian language in America. A study in bilingual behaviour*. Philadelphia: University of Pennsylvania Press.
- _____. 1956. *Bilingualism in the Americas. A bibliography and research guide*. Alabama: University of Alabama Press.
- _____. 1958. Language contact. *Proceedings of the Eighth International Congress of Linguists*, ed. by Eva Sivertsen, 771-85. Oslo: Oslo University Press.
- _____. 1973. Bilingualism, language contact, and immigrant languages in the United States: A research report 1956-1970. *Current Trends in Linguistics X. Linguistics in North America*, ed. by T.A. Sebeok, 505-91. The Hague: Mouton.
- HAUST, DELIA. 1993. Formen und Funktionen des Codeswitching. *Linguistische Berichte* 144.93-129.
- HEATH, JEFFREY. 1986. Hasta la mujerra! and other instances of playful language mixing in Morocco. *Mediterranean Language Review* 2.113-6.
- _____. 1987. *Ablaut and ambiguity. Phonology of a Moroccan Arabic dialect*. Albany, NY: State University of New York Press.
- _____. 1989. *From code-switching to borrowing. A case study of Moroccan Arabic*. London/New York: Kegan Paul International.
- HELLER, MONICA (Ed.) 1988. *Codeswitching. Anthropological and sociolinguistic perspectives*. Berlin etc.: Mouton de Gruyter.
- HERNÁNDEZ-CHAVEZ, EDUARDO, ANDREW D. COHEN, and ANTHONY F. BELTRAMO (Eds.) 1975. *El lenguaje de los Chicanos: regional and social characteristics used by Mexican Americans*. Arlington, VA: Center for Applied Linguistics.
- HERRERO MUÑOZ-COBO, BÁRBARA. 1996. *El árabe marroquí: aproximación sociolingüística*. Almería: Universidad de Almería.

- IGLA, BIRGIT. 1991. On the treatment of foreign verbs in Romani. *In the margin of Romani. Gypsy languages in contact*, ed. by Peter Bakker and Marcel Cortiade, 50-5. Amsterdam: Institute ATW, University of Amsterdam.
- JACOBSON, P. 1995. Constituent structure. *The Encyclopedia of Language and Linguistics*, ed. by R.E. Asher, vol. 2, 713-24. Oxford: Pergamon.
- JAKE, JANICE L. 1994. Intrasentential code switching and pronouns: on the categorial status of functional elements. *Linguistics* 32.271-98.
- _____ and CAROL MYERS-SCOTTON. 1997. Codeswitching and compromise strategies: implications for lexical structure. *The International Journal of Bilingualism* 1(1).25-39.
- JOSHI, ARAVIND K. 1981. Some problems in processing sentences with intrasentential code-switching. Paper presented at University of Texas Parsing Workshop, Austin.
- _____. 1982. Processing of sentences with intra-sentential code-switching. *COLING 82. Proceedings of the Ninth International Conference of Computational Linguistics, Prague, July 5-10, 1982*, ed. by Ján Horecký, 145-50. (North-Holland Linguistic Series 47.) Amsterdam etc.: North-Holland.
- _____. 1985. Processing of sentences with intrasentential code switching. *Natural language parsing. Psychological, computational, and theoretical perspectives*, ed. by David R. Dowty, Lauri Karttunen and Arnold M. Zwicky, 190-205. Cambridge: Cambridge University Press. [Revised version of Joshi 1982]
- KAMWANGAMALU, NKONKO M. 1989. Some morpho-syntactic aspects of French/English-Bantu code-mixing: Evidence for universal constraints. *Papers from the 25th annual regional meeting of the Chicago Linguistic Society. Part Two: Parasession on language in context*, ed. by Bradley Music, Randolph Graczyk and Caroline Wiltshire, 157-70. Chicago, IL: Chicago Linguistic Society.
- KAUFMAN, DORIT and MARK ARONOFF. 1989. Morphological interaction between L1 and L2 in language attrition. *Variation in second language acquisition. Volume II: Psycholinguistic issues*, ed. by Susan Gass, Carolyn Madden, Dennis Preston and Larry Selinker, 202-15. (Multilingual Matters 50.) Clevedon and Philadelphia: Multilingual Matters.
- KEDDAD, SADIKA. 1986. An analysis of French-Arabic code-switching in Algiers. PhD Thesis, University of London, Birbeck College.
- KEEGAN, JOHN M. 1984. Geminate clusters and Moroccan Arabic syllable structure. *CUNYForum Papers in Linguistics* 10.44-69.
- _____. 1986. The phonology and morphology of Moroccan Arabic. PhD Thesis, City University of New York.
- KHUBCHANDANI, LACHMAN M. 1968. The gender of English loan words in Sindhi. In Krishnamurti, 180-88.
- KLAVANS, JUDITH L. 1985. The syntax of code-switching: Spanish and English. *Selected papers from the XIIIth Linguistic Symposium on Romance Languages, Chapel Hill, NC, 24-26 March 1983* (Current Issues in Linguistic

- Theory, Vol. 36), ed. by Larry D. King and Catherine A. Maley, 213-31. Chapel Hill, CA: University of North Carolina.
- KRISHNAMURTI, BHADRIRAJU (ed.) 1968. *Studies in Indian linguistics. Professor M.B. Emeneau sastipurti volume*. Poona: Deccan College.
- KUHBERG, H. 1992. Longitudinal L2-attribution versus L2-acquisition in three Turkish children. *Second Language Research* 8.138-54.
- KULICK, DON 1992. *Language shift and cultural reproduction*. (Studies in the Social and Cultural Foundations of Language 14.) Cambridge: Cambridge University Press. [1977 reprint]
- _____ and CHRISTOPHER STROUD. 1990. Code-switching in Gapun: Social and linguistic aspects of language use in a language shifting community. *Melanesian Pidgin and Tok Pisin* (Studies in Language Companion Series 20), ed. by John W.M. Verhaar, 205-34. Amsterdam: John Benjamins.
- KURYŁOWICZ, JERZY. 1965. The evolution of grammatical categories. *Diogenes* 51.55-71.
- KWAN-TERRY, ANNA. 1992. Code-switching and code-mixing: The case of a child learning English and Chinese simultaneously. *Journal of Multilingual and Multicultural Development* 13(3).243-59.
- LAGHAOUT, MOHAMED. 1995. L'espace dialectal marocain, sa structure actuelle et son évolution récente. *Dialectologie et sciences humaines au Maroc*, 9-42. Rabat: Université Mohamed V, Faculté des Lettres et des Sciences Humaines.
- LAHLOU, MONCEF. 1991. A morphosyntactic study of code-switching between Moroccan Arabic and French. PhD Thesis, University of Texas at Austin.
- LALAMI, LAÏLA. 1996. Clitic left-dislocation in Moroccan Arabic. In Eid and Parkinson, 115-29.
- LEFEBVRE, CLAIRE. 1984. Grammaires en contact - définition et perspectives de recherche. *Revue québécoise de linguistique* 14(1).11-47.
- LEHTINEN, MERI. 1966. An analysis of a Finnish-English bilingual corpus. PhD Thesis, Indiana University.
- LEVELT, WILLEM J.M. 1989. *Speaking: From intention to articulation*. Cambridge, MA/London: MIT Press.
- LIPSKI, JOHN M. 1978. Code-switching and the problem of bilingual competence. *Aspects of bilingualism*, ed. by Michel Paradis, 250-64. Columbia, SC: Hornbeam Press.
- MAHOOTIAN, SHAHRZAD. 1993. A null theory of codeswitching. PhD Thesis, Northwestern University.
- _____ and BEATRICE SANTORINI. 1996. Codeswitching and the complement/adjunct distinction. *Linguistic Inquiry* 27(3).464-79.
- MARÇAIS, PH. 1957. Les parlers arabes. *Initiation à l'Algérie*, 215-37. Paris: Adrien Maisonneuve.
- MARTENS, E., J. ROIJEN and J. VEENMAN. 1994. *Minderheden in Nederland. Statistisch Vademecum 1993/1994*. Den Haag: SDU/CBS.

- MASHLER, YAEL. 1995. Structural and pragmatic contrasts shaping an emergent bilingual grammar. In *Summer School Code-Switching and Language Contact*, 142-52.
- MAZZON, GABRIELLA. 1993. English in Malta. *English Worldwide* 14(2).171-208.
- MCCLURE, ERICA. 1977. Aspects of code-switching in the discourse of Mexican-American children. *Linguistics and anthropology: Georgetown Round Table on Languages and Linguistics 1977*, ed. by Muriel Saville-Troike, 93-115. Washington, DC: Georgetown University Press.
- MEECHAN, MARJORIE and SHANA POPLACK. 1995. Orphan categories in bilingual discourse. A comparative study of adjectivization strategies in Wolof/French and Fongbe/French. In *Summer School Code-Switching and Language Contact*, 179-204.
- MERRIT, MARILYN, AILIE CLEGHORN, JARED O. ABAGI and GRACE BUNYI. 1992. Socialising multilingualism: Determinants of codeswitching in Kenyan primary classrooms. *Journal of Multilingual and Multicultural Development*, 13(1&2).103-21.
- MIFSUD, MANWEL. 1992. The non-Semitic element in Maltese verbal morphology. A descriptive and comparative study of loanverbs in Maltese. PhD Thesis, University of Malta.
- . n.d. [1995]. The productivity of Arabic in Maltese. *Proceedings of the 2nd international conference of L'association internationale pour la dialectologie arabe*, 151-60. [Cambridge]: [University of Cambridge].
- MILLER, ANN M. 1987. Phonetic characteristics of Levantine Arabic geminates with differing morpheme and syllable structures. *Working Papers in Linguistics (Ohio State University)* 36.120-40.
- MILROY, LESLEY and PIETER MUYSKEN (eds.) *One speaker, two languages. Cross-disciplinary perspectives on code-switching*. Cambridge: Cambridge University Press.
- MORAVCSIK, EDITH A. 1978. Universals of language contact. *Universals of human language. Vol I: Method and theory*, ed. by Joseph H. Greenberg, 94-122. Stanford, CA: Stanford University Press.
- MOYER, MELISSA. 1995. On defining matrix language in sentential code-switching: a syntactic approach. In *Summer School Code-Switching and Language Contact*, 192-204.
- MUUS, PHILIP. 1993. *Internationale migratie naar Europa. Een analyse van internationale migratie, migratiebeleid en mogelijkheden tot sturing van immigratie, met aandacht voor de Europese gemeenschap en Nederland*. Amsterdam: Sua.
- MUYSKEN, PIETER. 1981. Halfway between Quechua and Spanish. The case for relexification. *Historicity and variation in Creole studies*, ed. by Arnold Highfield and Albert Valdman, 52-78. Ann Arbor, MI: Karoma.

- _____. 1987. Neutrality in code-mixing. *Eigen en vreemd. Handelingen van het 39e Filologencongres Amsterdam*, 359-73. Amsterdam: Vrije Universiteit Amsterdam.
- _____. 1991. Needed: a comparative approach. In *Papers for the symposium on code-switching in bilingual studies*, vol. 1, 253-72. Strasbourg: ESF.
- _____. 1995. Code-switching and grammatical theory. In Milroy and Muysken, 177-98.
- _____ and VINCENT DE ROOIJ. 1995. Review of Myers-Scotton 1993a and 1993b. *Linguistics* 33(5).1043-50.
- MYERS-SCOTTON, CAROL. 1993a. *Social motivations for codeswitching. Evidence from Africa*. Oxford: Clarendon.
- _____. 1993b. *Duelling languages. Grammatical structure in codeswitching*. Oxford: Clarendon.
- _____. 1995. Language processing and the mental lexicon in bilinguals. *Current approaches to the lexicon. A selection of papers presented at the 18th LAUD symposium, Duisburg, March 1993*, ed. by René Dirven and Johan Vanparys, 73-100. Frankfurt: Peter Lang.
- _____. 1997. Code-switching. *The handbook of sociolinguistics*, ed. by Florian Coulmas, 217-37. Oxford: Blackwell.
- _____ and SHOJI AZUMA. 1990. A frame-based process model of codeswitching. *Papers from the 26th regional meeting of the Chicago Linguistic Society*, ed. by Michael Ziolkowski, Manuela Noske, and Karen Deaton, vol. 1, 307-21. Chicago, IL: Chicago Linguistic Society.
- _____ and JANICE L. JAKE. 1995. Matching lemmas in a bilingual language competence and production model. *Linguistics* 33(3).981-1024.
- _____, JANICE L. JAKE and MAHA OKASHA. 1996. Arabic and Constraints on Codeswitching. In Eid and Parkinson, 9-43.
- NAIT M'BAREK, MOHAMED, and DAVID SANKOFF. 1988. Le discours mixte arabe/français: emprunts ou alternances de langue? *Canadian Journal of Linguistics* 33(2).143-154.
- NARTEY, JONAS N.A. 1982. Code-switching, interference, or faddism? Language use among educated Ghanaians. *Anthropological Linguistics* 24.183-92.
- NISHIMURA, MIWA. 1983. On intrasentential code-switching in Japanese/English. *Proceedings of the XIIIth international congress of linguists, August 29 - September 4, 1982, Tokyo*, ed. by Shirô Hattori and Kazuko Inoue, 828-31. The Hague: Proc. Publishing Committee.
- _____. 1986. Intrasentential code-switching: The case of language assignment. *Language processing in bilinguals: psycholinguistic and neuropsychological perspectives*, ed. by Jyotsna Vaid, 123-43. Hillsdale, NJ/London: Lawrence Erlbaum.
- _____. 1989. The topic-comment construction in Japanese-English code-switching. *World Englishes* 8. 365-77.

- _____. 1995a. The use of Japanese topics in bilingual discourse in Japanese and English. *The twenty-first LACUS forum 1994*, ed. by Mava Jo Powell, 39-47. Chapel Hill, NC: LACUS (Linguistic Association of Canada and the US).
- _____. 1995b. A functional analysis of Japanese-English code-switching. *Journal of Pragmatics*. 23(2).157-81.
- NORTIER, JACOMINE. 1990. *Dutch-Moroccan Arabic Code Switching among Moroccans in the Netherlands*. Dordrecht: Foris.
- _____. 1992. Code-Switching in Moroccan Arabic/Dutch vs. Moroccan Arabic/French Language Contact. *OTS Working Papers* (OTS-WP-TL-92-015). Utrecht: Research Institute for Language and Speech.
- _____ and HENRIETTE SCHATZ. 1988. Van éénwoordwisseling naar ontlening, een vergelijkend onderzoek. *Forum der Letteren* 29.161-78.
- NUIJTENS, E.T.G. 1962. *De tweetalige mens. Een taalsociologisch onderzoek naar het gebruik van dialect en cultuurtaal in Borne*. Assen: Van Gorcum.
- OBDEIJN, HERMAN. 1994. Hier "travailleurs invités", aujourd'hui citoyens. In Basfao and Taarji, 489-91.
- OBRECHT, D.H. 1965. Three experiments in the perception of geminate consonants in Arabic. *Language and Speech* 8.31-41.
- OTTEN, ROEL. 1983. *Basiswoordenboek van het Marokkaans Arabisch: Marokkaans/Nederlands, Nederlands/Marokkaans*. Muiderberg: Coutinho.
- OUHALLA, JAMAL. 1991. *Functional categories and parametric variation*. London: Routledge.
- OZÓG, A.C.K. 1987. The syntax of the mixed language of Malay-English bilinguals. *RELC Journal*, 18(1).72-90.
- PANDIT, IRA. 1986. *Hindi English code switching - mixed Hindi English*. Delhi: Datta Book Centre.
- _____. 1990. Grammaticality in code switching. *Codeswitching as a worldwide phenomenon*, ed. by Rodolfo Jacobson, 33-69. New York, NY: Peter Lang.
- PAP, LEO. 1949. *Portuguese-American speech*. New York: King's Crown Press.
- PAPADEMETRE, LEO. 1994. Code interaction and communicative resourcing. In Philippaki-Warbuton et al., 349-56.
- Papers for the Symposium on Code-Switching in Bilingual Studies: Theory, Significance, and Perspectives, Barcelona, March 21-23, 1991*. Strasbourg: ESF.
- PETERSEN, JENNIFER. 1988. Word-internal code-switching constraints in a bilingual child's grammar. *Linguistics* 26(3).479-93.
- PFAFF, CAROL W. 1976. Functional and structural constraints on syntactic variation in code-switching. *Papers from the parasession on diachronic syntax, April 22, 1976*, ed. by Sanford B. Steever, Carol A. Walker, and Salikoko S. Mufwene, 248-59. Chicago, IL: Chicago Linguistic Society.
- _____. 1979. Constraints on language mixing: intrasentential code-switching and borrowing in Spanish/English. *Language* 55.291-318.

- PHILIPPAKI-WARBURTON, IRENE, KATERINA NICOLAIDIS and MARIA SIFIANOU (Eds.) 1994. *Themes in Greek Linguistics. Papers from the First International Conference on Greek Linguistics, Reading, September 1993*. (Current Issues in Linguistic Theory 117.) Amsterdam/Philadelphia: John Benjamins.
- POPLACK, SHANA. 1980. Sometimes I'll start a sentence in English y terminó en español: towards a typology of code-switching. *Linguistics* 18.581-618.
- _____. 1981. Syntactic structure and social function of codeswitching. In Durán, 169-84.
- _____. 1990. Variation theory and language contact: concepts, methods and data. *Papers for the workshop on concepts, methodology and data. ESF network on code-switching and language contact*, 33-66. Strasbourg: European Science Foundation.
- _____ and MARJORY MEECHAN. 1995. Patterns of language mixture: nominal structure in Wolof-French and Fongbe-French bilingual discourse. In Milroy and Muysken, 199-232.
- _____ and DAVID SANKOFF. 1988. Codeswitching. *Sociolinguistics. Soziolinguistik. An international handbook of the science of language and society*, ed. by U. Ammon, N. Dittmar and K. Mattheier, vol II, 1174-80. Berlin: De Gruyter.
- _____, DAVID SANKOFF and CHRISTOPHER MILLER. 1988. The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics* 26.47-104.
- _____, SUSAN WHEELER and ANNELI WESTWOOD. 1989. Distinguishing language contact phenomena: Evidence from Finnish-English bilingualism. *Bilingualism across the lifespan. Aspects of acquisition, maturity, and loss*, ed. by Kenneth Hyldenstam and Loraine K. Obler, 132-54. [first appeared in 1987 in: *The Nordic languages and modern linguistics* 6, ed. by P. Lilius and M. Saari, 33-56. Helsinki: University of Helsinki Press.]
- PROCHÁZKA, STEPHAN. n.d. [1995]. The Arabic dialects of Cilicia (Southern Turkey). *Proceedings of the 2nd international conference of L'association internationale pour la dialectologie arabe*, 189-96. [Cambridge]: [University of Cambridge].
- PUJOL, MERCE. 1991. L'alternance de langue comme signe de différenciation générationnelle. *Langage et société* 58.37-64.
- RHAÏB, DRISS. 1996. Le mélange de langues dans le discours d'enseignants marocains de français au Maroc, en dehors de la classe. *Revue québécoise de linguistique* 24(2).151-63.
- ROMAINE, SUSAN. 1989. *Bilingualism*. Oxford: Blackwell. [1995 edition]
- ROSETTI, AL. 1945/49. Langue mixte et mélange des langues. *Acta Linguistica* 5.73-9.
- ROTTENBERG, MICHEL. 1937. L'argot franco-turc du lycée de Galatasaray. *Le français moderne* 5.161-70.
- ROUCHDY, ALEYA (ed.) 1992. *The Arabic language in America*. Detroit: Wayne State UP.

- RUDIN, CATHERIN, and ALI EMINOV. 1990. Bulgarian Turkish: the linguistic effects of recent nationality policy. *Anthropological Linguistics* 32(1-2). 149-62.
- SAFI, SABAH. 1992. Functions of code-switching: Saudi Arabic in the United States. In Rouchdy, 72-80.
- SALMONS, JOE. 1990. Bilingual discourse marking: code switching, borrowing, and convergence in some German-American dialects. *Linguistics* 28. 453-80.
- SANKOFF, DAVID and SHANA POPLACK. 1981. A formal grammar for code-switching. *Papers in Linguistics: International Journal of Human Communication* 14(1).3-45.
- _____, SHANA POPLACK and SWATHI VANNIARAJAN, .1991. The empirical study of code-switching. In *Papers for the symposium on code-switching in bilingual studies: theory, significance, and perspectives*, vol. 1, 181-206.
- SANTORINI, BEATRICE, and SHAHRZAD MAHOOTIAN. 1995. Codeswitching and the syntactic status of adnominal adjectives. *Lingua* 96.1-27.
- SAVIĆ, JELENA M. 1995. Structural convergence and language change: Evidence from Serbian-English code-switching. *Language in Society* 24(4).475-92.
- SCHACHTER, PAUL. 1985. Parts-of-speech systems. In Shopen, 3-61.
- _____, and FE T. OTANES. 1972. *Tagalog reference grammar*. Berkeley etc.: University of California Press. [1983 reprint]
- SCHATZ, HENRIETTE F. 1989. Code-switching or borrowing? English elements in the Dutch of Dutch-American immigrants. *ITL Review of Applied Linguistics* 83.125-62.
- SCHIFFRIN, DEBORAH. 1987. *Discourse markers*. Cambridge: Cambridge University Press.
- SCHMID, BEATA. 1986. Constraints on code-switching: Evidence from Swedish and English. *Nordic Journal of Linguistics* 9(1).55-82.
- SCHUCHARDT, HUGO. 1884. *Dem Herrn Franz von Miklosich zum 20. Nov. 1883. Slawo-deutsches und Slawo-italienisches*. Graz: Leuschner and Lubensky.
- SCOTTON, CAROL M. 1988. Code-switching and types of multilingual communities. *Language spread and language policy*, ed. by Peter H. Lowenberg, 61-82. Washington, DC: Georgetown University Press.
- SEAMAN, P. DAVID. 1972. *Modern Greek and American English in contact*. The Hague/Paris: Mouton.
- SHANMUGAN PILLAI, M. 1968. English borrowings in educated Tamil. In Krishnamurti, 296-306.
- SHOPEN, TIMOTHY (Ed.) 1985 *Language Typology and Syntactic Description. Vol. I Clause structure*. Cambridge: Cambridge University Press.
- SINGH, RAJENDRA. 1995. Code-switching/mixing: patterns and bilingual proficiency. In *Summer School Code-Switching and Language Contact*, 142-52.
- SLAOU, HADIA. 1986. Problèmes de transitions de langues posés par le discours bilingue Arabe-Marocain/Français. Thèse de Doctorat de 3ème cycle, Université de Paris VII.

- SMART, J.R. 1990. Pidginization in Gulf Arabic: a first report. *Anthropological Linguistics* 32(1-2).83-119.
- SOBOLEWSKI, FRANK A. 1982. Some syntactic constraints in Tagalog-English language mixing. *Philippine Journal of Linguistics*, 13(2).35-62.
- SOUNKALO, J. 1995. Code-switching as indexical of native language lexical deficiency in Mauritania. *Journal of Multilingual and Multicultural Development* 16(5):403-422.
- SRIDHAR, S.N., and KAMAL K. SRIDHAR. 1980. The syntax and psycholinguistics of bilingual code-mixing. *Canadian Journal of Psychology* 34.407-16.
- STENSON, NANCY. 1990. Phrase structure congruence, government, and Irish-English code-switching. *The syntax of the modern Celtic languages* (Syntax and Semantics 23), ed. by Randall Hendrick, 137-97. San Diego, CA: Academic Press.
- STUBBS, MICHAEL. 1983. *Discourse analysis: The sociolinguistic analysis of natural language*. Oxford: Blackwell.
- _____. 1994. *Summer School Code-Switching and Language Contact. Papers from the Summer School Code-Switching and Language Contact, Ljouwert/Leeuwarden, 14-17 September 1994*, Ljouwert/Leeuwarden: Fryske Akademy.
- TAHITU, EGBERTUS. 1989. Melaju Sini. Het Maleis van Molukse jongeren in Nederland. PhD Thesis, University of Leiden.
- THOMASON, SARAH G., and TERENCE KAUFMAN. 1988. *Language contact, creolization and genetic linguistics*. Berkeley etc.: University of California Press.
- TIMM, LENORA A. 1975. Spanish-English code-switching: el porqué y how-not-to. *Romance Philology* 28(4).473-82.
- _____. 1994. The limits of code-switching constraints, with some evidence from Breton/French switching. *Journal of Celtic Linguistics* 3.95-134.
- TORRES, LOURDES. 1989. Code-mixing and borrowing in a New York Puerto Rican community: a cross-generational study. *World Englishes* 8(3).419-32.
- TREFFERS-DALLER, JANINE. 1994. *Mixing two languages. French-Dutch contact in a comparative perspective* (Topics in Sociolinguistics 9). Berlin/New York: Mouton de Gruyter.
- _____. 1995. Code-switching between a case-assigner and its complement: the case of Turkish-German prepositional phrases. In *Summer School Code-Switching and Language Contact*, 253-63.
- VAN DER MEER, PH. 1984. *Omgaan met Marokkanen*. Deventer: Van Loghum Slaterus.
- VAN NESS, SILKE. 1994. Die Dimensionen lexikalischer Entlehnungen im Pennsylvaniendeutschen von Ohio (USA). *Zeitschrift für Dialektologie und Linguistik* 61(3).279-97.
- VAN STADEN, MIRIAM. 1997. Borrowing and code-switching: Where does Tidorese end and Malay begin? Paper presented at Perspectives on the Bird's Head of Irian Jaya, Indonesia. Leiden, The Netherlands, 13-17 October 1997.

- _____. (forthcoming). Where does Malay end and Tidore begin? *Proceedings of the Perspectives on the Bird's Head Conference*, ed. by Jelle Miedema, Cecilia Odé and Rien Dam, 00-00.
- VEENMAN, JUSTUS. 1994a. Les Marocains sur le marché du travail. In Basfao and Taarji, 509-11.
- _____. 1994b. Instruction: profil général. In Basfao and Taarji, 512-3.
- VOCKE, S., and W. WALDNER. 1982. *Der Wortschatz des anatolischen Arabisch*. Erlangen: Universität Erlangen.
- VOIGT, HERMAN A. 1994. Code-wisseling, taalverschuiving en taalverandering in het Melaju Sini. Master's Thesis, University of Tilburg.
- _____. 1997. Taalvervreemding in het Melaju Sini. *Toegepaste Taalwetenschap in Artikelen* 57(2).67-79.
- WAGNER, DANIEL A. 1993. *Literacy, culture and development. Becoming literate in Morocco*. Cambridge: Cambridge University Press.
- WARD, C. 1975. The Serbian and Croatian communities in Milwaukee. *General Linguistics*, 16(2-3).151-65.
- WEINREICH, URIEL. 1953. *Languages in contact, findings and problems*. New York, NY: Linguistic Circle of New York.
- WENTZ, JAMES PAUL. 1977. Some considerations in the development of a syntactic description of code-switching. PhD Thesis, University of Illinois.
- _____. and ERICA MCCLURE. 1977. Monolingual "codes". Some remarks on the similarities between bilingual and monolingual code-switching. *Papers from the 13th regional meeting, Chicago Linguistic Society, April 1977*, ed. by Woodford A. Beach, Samuel E. Fox, and Shulamith Philosoph, 706-13. Chicago, IL: Chicago Linguistic Society.
- WERDMÖLDER, HANS. 1994. Délinquance des jeunes: le cumul des handicaps. In Basfao and Taarji, 544-46.
- WERNITZ, CORINNA J. 1994. *Bedingungen und Voraussetzungen für Sprachwechsel. Eine Untersuchung zum Sprachwechsel bei bilingualen Marokkanern in Frankreich*. Frankfurt/M.: Peter Lang.
- WETZER, HARRIE. 1995. Nouniness and Verbiness. A typological study of adjectival predication. PhD Thesis, University of Nijmegen.
- WOOLFORD, ELLEN. 1984. On the application of Wh-movement and inversion in codeswitching sentences. *Revue québécoise de linguistique* 14(1).77-86.
- YOUSSE, ABDERRAHIM. 1992. *Grammaire et lexique de l'arabe marocain moderne*. Casablanca: Wallada.
- ZENTELLA, ANA CELIA. 1981. Tá bien, you could answer me en cualquier idioma: Puerto Rican codeswitching in bilingual classrooms. In Durán, 109-31.

Appendix

Sociolinguistic data concerning the respondents of the Nijmegen corpus were noted down on a form by Youssef Azghari during 1990-1991. This form, devised by Jacomine Nortier, is reproduced below.

SOCIOLINGUISTISCH PROFIEL

sociolinguistic profile

Niet invullen waar de informant bij is. Schrijf zo veel mogelijk op!!
don't fill it out in the presence of the informant. write down as much as you can.

1. Naam van de informant en pseudonym; de echte naam wordt nergens anders gebruikt.
informant's name and pseudonym; the real name will not be used elsewhere.
2. Nummer van de geluidscassette(s), datum van de opname, duur, locatie en situatie.
number of the audio tape, date of recording, duration, location and situation.
3. Geslacht:
sex
4. Leeftijd en geboortjaar:
age and year of birth
5. Geboorteplaats en land:
place and country of birth
6. Hoeveel jaren waar gewoond?
how many years lived where
7. Opleiding en evt. beroep (soort en aantal jaren) en evt. toekomstplannen (waar?)
education and profession, if applicable, (kind and number of years) and future plans, if any (in which country?)
8. Totaal aantal jaren opleiding? Welke instructieta(a)l(en) ?

- total number of years of education? which languages of instruction?
9. Woont hij/zij met familie? Plaats daarin?
does he/she live with her/his family? place in the family?
 10. Welke nationaliteit(en) hebben zijn/haar vrienden?
which nationality (-ies) do his/her friends have?
 11. Vrijtijdsbesteding, hobbies:
spare time activities, hobbies
 12. Heeft hij/zij andere talen geleerd en zoja, waar?
did he/she acquire other languages and if so, where?
 13. Welke ta(a)l(en) spreekt hij/zij het best naar EIGEN OORDEEL? En hoe zit het met de andere ta(a)l(en)?
which language(s) does he/she speak best ACCORDING TO HIM/HERSELF? what about the other language(s)?
 14. Welke taal spreekt hij/zij het liefst? En de andere ta(a)l(en)?
which language does he/she prefer to speak? what about the other language(s)?
 15. Naar welk soort televisie- en radioprogramma's kijkt of luistert hij/zij graag en vooral: in welke taal?
which TV and radio programmes does he/she like to watch or listen to, and in particular: in which language?
 16. Welke ta(a)l(en) spreekt hij/zij met ouder(s), kind(eren), broer(s), zus(sen), overige familieleden, vrienden, collega's, klasgenoten etc.?
(per groep afzonderlijk beantwoorden, svp)
which language(s) does he/she speak with parent(s), child(ren), brother(s), sister(s), other family members, friends, colleagues, class mates, etc.? (please answer for each group separately)
 17. In welke ta(a)l(en) leest en schrijft de informant? Denk aan krant(en), boeken, agenda, tellen, rekenen en misschien zelfs denken.
in which language(s) does the informant read and write? think of newspaper(s), books, agenda, counting, calculating and maybe even thinking.
 18. Eerste indruk van de persoonlijkheid (verlegen, kletschous, coöperatief).
first impression of the personality (shy, talkative, cooperative).
 19. Overige opmerkingen:
further remarks.

Samenvatting

Dit proefschrift gaat over codeswitching in het algemeen en Marokkaans Arabisch/Nederlandse codeswitching in het bijzonder. Codeswitching is het afwisselend gebruik van twee of meerdere talen in hetzelfde gesprek, variërend van het gebruik van losse woorden uit de ene taal in de context van een andere taal tot de afwisseling van grotere stukken eentalige tekst in verschillende talen. Deze studie beperkt zich tot de morfologische en syntactische kenmerken van codeswitching.

Het doel van dit proefschrift is tweeledig. Ten eerste wil het een bijdrage leveren aan de discussie over hoe de morfo-syntactische eigenschappen van codeswitching het best beschreven en geïnterpreteerd kunnen worden; ten tweede geeft het een gedetailleerde beschrijving van een corpus van tweetalige gesprekken in het Marokkaans Arabisch en het Nederlands.

De opzet van dit boek is als volgt. De elf hoofdstukken zijn over drie delen verdeeld. Deel I (Hoofdstuk 1 tot en met 3) benadert codeswitching vanuit een theoretisch en taalvergelijkend perspectief en doet een voorstel voor een beschrijvings- en verklaringsmodel. Deel II (Hoofdstuk 4 tot en met 9) beschrijft het corpus Marokkaans Arabisch/Nederlandse gesprekken dat in Nijmegen is verzameld. Deel III (Hoofdstuk 10 en 11) evalueert het in Deel I voorgestelde model in het licht van de in Deel II gepresenteerde gegevens en bespreekt mogelijkheden voor verder onderzoek naar een volgend niveau van verklaring voor de regelmatigheden in codeswitching-gedrag.

Hoofdstuk 1 geeft een overzicht van de literatuur over codeswitching, met nadruk op de studie van de grammaticale aspecten ervan. De grammaticale studies worden onderverdeeld in een lineaire en een structurele benadering. De lineaire benadering ziet codeswitching als een van links naar rechts verlopend proces, tussen een element x uit de ene taal en een element y uit de andere taal. Deze benadering bestudeert de mogelijke juxtaposities van elementen uit beide talen en formuleert beperkingen daarop in de trant van 'er kan geen switch zijn tussen een Spaans Subject pronomen en een Engels finiet werkwoord'. De beperkingen op zulke juxtaposities worden verklaard uit woordvolgorde-verschillen tussen de twee betrokken talen, of door een aparte grammatica voor de codeswitching-variëteit aan te nemen. De structurele benaderingen gaan uit van hiërarchische relaties tussen morfemen, woorden en zinsdelen en verklaren de regelmatigheden in codeswitching-gedrag vanuit deze relaties. De structurele benaderingen worden weer onderverdeeld in *government*- en insertie-modellen. De *government* of 'regeer-' modellen kennen een structurerende functie toe aan bepaalde morfeem- of woordcategorieën. De insertiemodellen gaan

ervan uit dat één taal, de matrixtaal, de morfo-syntactische structuur levert waarin elementen uit de andere taal kunnen worden ingebed; die andere taal wordt hier de ingebedde taal genoemd. De voor- en nadelen van deze verschillende benaderingen worden in Hoofdstuk 1 besproken, waarbij een voorkeur voor het insertiemodel naar voren komt. De centrale kwestie die bij insertiemodellen speelt, is hoe de matrixtaal moet worden gedefinieerd en, daarmee samenhangend, wat voor elementen uit de andere taal kunnen worden ingebed. Deze kwestie is in de bestaande insertie-modellen nog niet bevredigend opgelost.

In Hoofdstuk 2 wordt voortbouwend op eerdere voorstellen een insertiemodel ontwikkeld dat de *Monolingual Structure Approach* (MSA) wordt genoemd. De MSA onderscheidt zich door de volgende kenmerken. Ten eerste de definitie van de matrixtaal: Deze wordt uitsluitend bepaald op grond van morfologische en syntactische eigenschappen van de matrix-structuur zelf, waarbij zowel de finiete zin als de zinsconstituent een matrixstructuur kan zijn. Teneinde in een later stadium insertiepatronen te kunnen relateren aan de sociale status van de matrixtaal en aan andere sociolinguïstische variabelen moet de matrixtaal onafhankelijk van dergelijke overwegingen herkend worden. De matrixtaal wordt op twee niveau's verschillend gedefinieerd:

Op het niveau van de zinsconstituent wordt de matrixtaal afgeleid uit de interne structuur van deze constituent: de matrixtaal is de taal waaraan de distributie van alle morfemen in de constituent kan worden toegeschreven. Met distributie wordt bedoeld zowel het vóórkomen van het morfeem als zijn volgorde ten opzichte van de andere morfemen in de constituent. Dit betekent dat de matrixtaal in eerste instantie ad hoc wordt bepaald, maar op grond van generalisaties kan de matrixtaal toch voorspeld worden. Het principe van generalisatie is erop gericht het aantal typen insertie binnen een bepaalde set data te beperken. Zo komen we tot de conclusie dat inhoudsmorfemen en sommige complexere vormen van inhoudswoorden veelvuldig ingebed worden, terwijl bijvoorbeeld lidwoorden, demonstrativa en 'losse' affixen steeds tot de matrixtaal behoren. Bepaalde affixen kunnen wel als deel van een complexe woordvorm ingebed worden; een veel voorkomend voorbeeld is de insertie van meervoudsnomina.

De constituent die één en precies één finiet werkwoord bevat wordt de finiete zin of *finite clause* genoemd. Op dit niveau is het mogelijk om een onafhankelijk criterium voor de matrixtaal aan te wijzen. De matrixtaal op het niveau van de *finite clause* is de taal van de inflectie van het finiete werkwoord. Deze matrixtaal bepaalt de distributie van zinsconstituenten, dat wil zeggen het vóórkomen van iedere constituent en de volgorde van de constituenten ten opzichte van elkaar.

In een *finite clause* kunnen constituenten van een andere taal dan de matrixtaal worden ingebed; in een zinsconstituent kunnen constituenten van een lagere orde worden ingebed, alsmede woorden en morfemen. Morfemen worden hier beschouwd als constituenten op het laagste niveau. Een *finite clause* kan ook zelf een zinsconstituent in een andere *finite clause* zijn. De MSA staat toe dat op meer dan één niveau tegelijk insertie plaatsvindt: een ingebedde constituent kan zelf een matrix

zijn waarin weer een element uit de andere taal is ingebed. Het is bijvoorbeeld mogelijk dat een Franse finiete zin een Arabische nominale constituent bevat, terwijl deze nominale constituent zelf weer een Frans nomen bevat.

De MSA gaat ervan uit dat codeswitching in alle gevallen in termen van insertiepatronen kan worden beschreven. Dit wil zeggen dat in alle gevallen een matrixtaal kan worden geïdentificeerd die de distributie van elementen in de matrixstructuur kan verklaren. De MSA is niet alleen een beschrijvingsmodel; zij biedt ook een eerste niveau van verklaring door een groot deel van de grammaticale regelmatigheden aan de grammatica van de matrixtaal toe te schrijven, en de interne structuur van de ingebedde constituenten aan de grammatica van de ingebedde taal. Het matrixtaal-model wordt evenwel ondermijnd door een aantal steeds terugkerende typen van tegenvoorbeelden. De volgende typen tegenvoorbeelden worden herkend: 1) het voorkomen van redundante morfemen wanneer hetzelfde kenmerk in zowel de ene als de andere taal wordt uitgedrukt; 2) de soms stelselmatige afwezigheid van bepaalde functiemorfemen die op grond van de matrixtaal-grammatica vereist zijn, en 3) woordvolgorde-patronen die niet door de matrixtaal verklaard kunnen worden. Deze problemen worden erkend, maar tegelijkertijd wordt betoogd dat het insertiemodel het meest geschikt is om de codeswitching-patronen die het model ondermijnen te herkennen en te beschrijven. De tegenvoorbeelden zijn er niet in alle soorten en maten, maar komen in herkenbare patronen voor, en die zouden veel moeilijker te herkennen zijn in een benadering die juxtaposities inventariseert. Voor de verschillende typen tegenvoorbeelden moeten aparte verklaringen gezocht worden die soms kunnen aansluiten bij de verklaringen voor vergelijkbare verschijnselen in eentalig taalgebruik.

Hoofdstuk 3 verkent een aantal probleemgebieden voor de MSA, samengebracht onder de noemer tekstgrammatica. Met deze term worden lexicale elementen en grammaticale constructies aangeduid die functioneren op het niveau van de tekst als geheel, hetzij als markeerders van de interne structuur van die tekst (zoals paragrafen in een geschreven tekst), hetzij als markeerders van de houding van de sprekers ten opzichte van wat er gezegd wordt, of ten opzichte van hun gesprekspartners (subjectieve modaliteit). Verschillende aspecten van tekstgrammatica zijn moeilijk in een insertiemodel te vatten. Adverbia van subjectieve modaliteit lijken vaak de woordvolgorde-eigenschappen van hun brontaal te behouden wanneer ze als ingebedde elementen voorkomen. Hetzelfde doet zich overigens ook wel voor bij adverbia die objectieve modaliteit, aspect of tijd uitdrukken en, in een enkel taalpaar, bij modale werkwoorden die dus niet onder de noemer 'tekstgrammatica' vallen. Een ander probleem vormen conjuncties en (andere) partikels die *discourse markers* genoemd worden. Deze vallen syntactisch gezien buiten de *finite clause* en dus buiten de competentie van de MSA. Hetzelfde geldt voor bepaalde taalspecifieke vormen van vooropplaatsing als een manier om de tekst te structureren. Het is mogelijk deze verschijnselen in een insertiemodel op te nemen door ook boven het niveau van de *finite clause* een matrixstructuur aan te nemen, maar op dit niveau levert de formulering van een niet-circulaire definitie van de matrixtaal onoverkomelijke

problemen op. Over het algemeen kan de distributie van zulke discourse markers ten opzichte van de zin en in de tekst als geheel het beste aan de hand van de brontaal verklaard worden.

Deel II van het proefschrift past de in Hoofdstukken 2 en 3 ontwikkelde benadering toe op het in Nijmegen verzamelde corpus Marokkaans Arabisch/Nederlandse teksten. Hoofdstuk 4 geeft de achtergronden van de dataverzameling, van de individuele respondenten en van de Marokkaanse gemeenschap in Nederland. De Hoofdstukken 5 tot en met 8 bespreken verschillende typen van Nederlandse inserties in Arabische matrixen. In vergelijking met andere codeswitchingdata en met de voorspellingen van de MSA vallen een aantal kenmerken op. Ten eerste is er de tendens om voorafgaand aan ingebedde Nederlandse nomina het Arabische lidwoord *l-* en ook de prepositie *l-* weg te laten (Hoofdstuk 5). Ten tweede worden Nederlandse werkwoorden ingebed door middel van een perifrastische doe-constructie. In de idiolecten van sommige respondenten is het gebruik van het Marokkaans Arabische werkwoord *dar* “doen” in deze functie sterk gegrammaticaliseerd (Hoofdstuk 6). Ten derde is de insertie van Nederlandse complexere constituenten zeldzaam vergeleken met de insertie van inhoudswoorden. Hoofdstuk 9 bespreekt de Marokkaans Arabische inserties in Nederlandse finiete zinnen en constituenten. Deze inserties zijn gering in aantal vergeleken bij de Nederlandse en ze hebben ook andere eigenschappen. Grofweg worden drie typen onderscheiden: 1) de insertie van cultuur-specifieke Arabische nomina die Nederlandse lexicale leemten opvullen; 2) de insertie van Arabische nomina en adjectiva in herhalingen; en 3) de insertie van Arabische prepositionele, adverbiale en nominale constituenten, met name als eerste constituent van een Nederlandse zin. Arabische nominale constituenten in Nederlandse zinnen komen in absolute aantallen ongeveer evenveel voor als Nederlandse constituenten in Arabische zinnen. Naast insertie-typen worden in de Hoofdstukken 8 en 9 ook enkele *discourse markers* behandeld.

Hoofdstuk 10 in Deel III evalueert de MSA in het licht van de Marokkaans Arabisch/Nederlandse data in II. De MSA wordt aan de hand van een aantal voorbeelden vergeleken met Myers-Scotton's *Matrix Language Frame model*. De conclusie is dat de MSA beter in staat is om de volgorde van zinsconstituenten in de *finite clause* te voorspellen, terwijl ook de mogelijkheid van gelaagde inbedding in de MSA veel data kan verklaren die voor het *Matrix Language Frame model* problematisch blijven. Verder komen dezelfde problemen die in Hoofdstuk 2 waren besproken opnieuw aan de orde, aangevuld met de perifrastische doe-constructies besproken in Hoofdstuk 6 die vaak moeilijk in een insertie-model gevat kunnen worden. Deze gebieden blijven problematisch voor zowel de MSA als het *Matrix Language Frame model*, maar de conclusie is wederom dat de voordelen van de insertiebenadering opwegen tegen de nadelen die de beperkingen van deze benadering met zich meebrengen. Als alternatieve oplossing zou men zich een aparte grammatica, dat wil zeggen een aparte verzameling regels, kunnen voorstellen voor de codeswitching-variant. Hierbinnen zou men dan ook het gegrammaticaliseerde gebruik van de perifrastische doe-constructie kunnen opnemen. Blijft echter dat vooralsnog

het merendeel van de regels van deze codeswitching-variant het meest economisch beschreven wordt aan de hand van de regels van de eentalige varianten en de aanname van een matrixtaal.

Hoofdstuk 10 gaat verder in op het feit dat de MSA als insertie-benadering niet alle aspecten van codeswitching dekt, ook niet voor zover die als insertiepatronen beschreven kunnen worden. Het is namelijk zo dat van alle logisch denkbare insertiepatronen er maar een paar met regelmaat voorkomen. Zo zijn de Marokkaans Arabisch/Nederlandse insertiepatronen bijvoorbeeld sterk asymmetrisch: het grootste deel betreft de inbedding van Nederlandse inhoudswoorden. Binnen de categorie 'inbedding van Nederlandse inhoudswoorden' zijn bepaalde patronen weer opvallend afwezig: Er worden vrijwel geen losse Nederlandse attributieve adjectiva ingebed maar wel Nederlandse adjectief-nomen combinaties, terwijl ook ingebedde Nederlandse nomina vrijwel nooit gemodificeerd worden door een Arabisch (matrixtaal) adjectief. Vergelijkbare beperkingen op gemengde combinaties van het type 'lexicaal hoofd + modificeerder' doen zich voor met betrekking tot adverbia. Dit leidt tot de formulering van de zogenaamde 'modificatie-restrictie'. Iets dergelijks vinden we ook in het geval van ingebedde Nederlandse werkwoorden, die uitsluitend met Nederlandse lexicale complementen voorkomen en niet met Arabische; dit wordt aangeduid als de 'complementerings-restrictie'.

Hoofdstuk 11 gaat in op de verklaringen voor patronen van insertie. Twee hoofdtypen van verklaringen worden hierbij onderscheiden, te weten sociolinguïstische en psycholinguïstische. De eerste soort verklaringen relateert insertiepatronen aan sociale variabelen betreffende de taalcontactsituatie en de relatieve status van de betrokken talen. Psycholinguïstische verklaringen refereren aan het spraakproces in de individuele spreker en de interactie van lexicale eenheden en regels van beide talen. Bij wijze van verkenning worden enkele sociolinguïstische en procesmatige verklaringen voorgesteld.

De asymmetrische insertiepatronen in Marokkaans Arabisch/Nederlands worden kenmerkend geacht voor een sociolinguïstische setting waarin één taal als de groepstaal fungeert en de andere taal opgelegd is als de economische en/of cultureel dominante taal. Omwille van deze generalisatie wordt de dichotomie Groepstaal/Opgelegde Taal voorgesteld, die onafhankelijk is van de tegenstelling Matrixtaal/Ingebedde Taal. De Nederlandse inserties in het Marokkaans Arabisch zijn typisch voor inserties van de Opgelegde Taal; de Marokkaans Arabische inserties zijn typisch voor inserties van de Groepstaal. Deze typologie kan verfijnd worden door op een meer gedetailleerd niveau te kijken naar de correlaties tussen insertie-typen en typen van tweetalige gemeenschappen. Inserties van de Groepstaal zijn heel verschillend van inserties van de Opgelegde taal: Terwijl inserties van de Opgelegde Taal vooral bestaan uit (semantische specifieke) inhoudswoorden, zijn inserties van de Groepstaal vaak juist nominale en prepositionele constituenten. De ingebedde constituenten van de Groepstaal zijn bovendien vaak niet semantisch specifiek en bestaan daarentegen vaak uit functiemorfemen, eventueel aangevuld met ingebedde inhoudsmorfemen van de Opgelegde Taal. Dit suggereert dat de ingebedde constituenten van de Groepstaal

een soort voortzetting zijn op lokaal niveau van de functie die de Groepstaal normaliter heeft als de matrixtaal, die het morfo-syntactische kader levert waarin inhoudswoorden uit de Opgelegde Taal worden ingebed.

Een procesmatige verklaring wordt voorgesteld voor de ‘modificatie-restricties’ en de ‘complementerings-restrictie’ die werden signaleerd in Hoofdstuk 10. Het valt op dat deze restricties betrekking hebben op dezelfde formele relaties als de collocaties die vaak als ingebedde elementen voorkomen, namelijk de combinatie van adjectief en nomen en van lexicaal direct object en werkwoord. De verbanden tussen inhoudswoorden die ertoe leiden dat collocaties vaak in hun geheel worden ingebed zouden wel eens verantwoordelijk kunnen zijn voor de waargenomen restricties. Met andere woorden, de restricties zouden verklaard kunnen worden wanneer wordt aangenomen dat collocationale verbanden een voorwaarde of althans een belangrijke faciliterende factor zijn voor het vóórkomen van adjectief-nomen en lexicaal object-werkwoord combinaties. De herhaling van een inhoudswoord als retorisch stijlfiguur of omwille van de cohesie van de tekst zou een mechanisme kunnen zijn dat beperkingen op gemengde modificatie- en complementeringsrelaties opheft. Tenslotte wordt nader ingegaan op het belang van herhaling als mechanisme op zichzelf voor het veroorzaken van inhoudswoord-insertie, het verspreiden van codeswitching-patronen over verschillende sprekers en het veroorzaken van sommige minder gebruikelijke patronen.

CURRICULUM VITAE

Louis Boumans (Tunis, 1967) studied Arabic language and literature at the University of Leiden. During a stay at Mohamed V University in Rabat in 1988-1989 he carried out research for his MA thesis on Moroccan Islamic family law and acquired the Moroccan variety of spoken Arabic. He received his Master's degree in 1992. In 1993, at the University of Nijmegen, he began work on his PhD project which forms the subject of this thesis. In June 1998 he was appointed as a postgraduate researcher at the University of Leiden, where he pursues his interests in Moroccan Arabic within the framework of a large-scale interdisciplinary survey of the multi-ethnic neighbourhoods Lombok and Transvaal in Utrecht.